Detecting Psychiatric Morbidity After Stroke
Comparison of the GHQ and the HAD Scale

Suzanne O’Rourke, PhD; Siobhan MacHale, MRCPsych; David Signorini, MA; Martin Dennis, MD

Background and Purpose—Mood disorders are common after stroke and may impede physical, functional, and cognitive recovery, making early identification and treatment of potential importance. We aimed to compare the accuracy of the General Health Questionnaire (GHQ-30) and the Hospital Anxiety and Depression (HAD) Scale in detecting psychiatric morbidity after stroke and to determine the most suitable cutoff points for different purposes.

Methods—One hundred five hospital-referred stroke patients completed both the GHQ-30 and HAD Scale 6 months after onset before a blinded psychiatric assessment in which the Schedule for Affective Disorders and Schizophrenia with some supplementary questions was used to determine a DSM-IV (Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition) diagnosis. Measures were compared in terms of sensitivity, specificity, and receiver operating characteristic curves.

Results—No significant differences were found between the GHQ-30 and the HAD Scale in identifying those patients with any DSM-IV diagnosis (P=0.95), grouped depression (P=0.56), or anxiety (P=0.25) disorders. The previously recommended cutoff points for identifying “cases” for the GHQ (4/5) and for the HAD Scale (8/9 and 11/12) were found to be suboptimal in this population.

Conclusions—The GHQ-30 and HAD scale exhibited similar levels of sensitivity and specificity. Data are presented, taking into account the “cost” of false-positives and negatives, to allow a choice of cutoff points suitable for differing situations. (Stroke. 1998;29:980-985.)

Key Words: anxiety ■ depression ■ stroke

Each year 127 000 strokes are treated in English hospitals, with an incidence of 2.9 per 1000 persons in England and Wales.1 In the first year after stroke, mood disorders have been estimated to affect between 23% and 60% of patients,2–6 more than twice the proportion in the general elderly population7 or in populations matched for physical disability.5 Such figures are particularly significant because depression is thought to impede physical, functional, and cognitive recovery.9–11 If early treatment were shown to improve any of these aspects of recovery, early identification would clearly be important.

Ideally, the diagnosis of psychiatric illness is made by standardized psychiatric interview.12,13 However, this may be impractical when screening large numbers of patients for treatable psychiatric disease or in large-scale research studies. In these situations, self-report questionnaires may be useful. Although several have been used with stroke patients, there is still considerable uncertainty regarding their suitability and the optimum cutoff points for different uses. In clinical practice these measures may be useful as screening tools to identify patients in need of further intervention; however, in choosing suitable cutoff points, the consequences of both false-negatives (ie, patients not receiving treatment) and false-positives (ie, wasted resources) must be considered. In randomized trials in which psychological outcomes are important, the power of the study is reduced when outcomes are misclassified; thus, an outcome instrument with a high accuracy is essential.

The GHQ14 and the HAD Scale15 are among the most commonly used measures of psychiatric morbidity after stroke. The GHQ was designed as a screening instrument to identify psychiatric disorders. It does not aim to provide a diagnosis but rather to identify those in need of further psychiatric assessment. We specifically chose the 30-item version because items relating to physical illness have been removed to make it suitable for our physically ill population. Unlike the 28-question version, it is not split into subscales for depression and anxiety. We selected the HAD Scale for comparison because its authors had specifically attempted to improve on the GHQ.15 Substantially shorter than most versions of the GHQ, it gives each patient a score on one of its two subscales, anxiety and depression. The HAD Scale, which was designed for use in nonpsychiatric hospital clinics, specifically avoids contamination by excluding questions referring to physical complaints.15

Received December 17, 1997; final revision received January 29, 1998; accepted February 2, 1998.
From the University of Edinburgh (S.O’R., D.S., M.D.), Neurosciences Trials Unit, Department of Clinical Neurosciences, The University of Edinburgh, and the Edinburgh Health Care NHS Trust (S.M.), Royal Edinburgh Hospital, Edinburgh, Scotland.
Correspondence to Martin Dennis, MD, University of Edinburgh, Neurosciences Trials Unit, Department of Clinical Neurosciences, The University of Edinburgh, Bramwell Dott Building, Western General Hospital, Crewe Rd, Edinburgh, UK EH4 2XU.
E-mail MSD@skull.dcn.ed.ac.uk
© 1998 American Heart Association, Inc.
In this study we compared the accuracy of the GHQ-30 and HAD Scale in the identification of patients with current psychiatric morbidity assessed using a standardized semi psychiatric interview, the SADS, with supplementary questions to generate a DSM-IV diagnosis.

Subjects and Methods

We identified stroke patients as part of a randomized trial of a "stroke family care worker," in which patients referred to our hospital with a clinical diagnosis of stroke, confirmed by CT scan, were entered. Criteria for inclusion in the randomized trial were assessment at the hospital within 1 month of stroke onset, residence within 25 miles, consent to follow-up, likelihood of survival, and acute stroke as the dominant illness.

A psychology research associate (S.O’R.) visited patients in their own homes 6 months after randomization and, as part of an extensive test battery, administered the GHQ-30. The GHQ-30 is specifically concerned with the subjects’ health in the previous “few” weeks, specifying that “we want to know about present and recent complaints, not those you had in the past.” The response options typically include “not at all,” “no more than usual,” “rather more than usual,” and “much more than usual.” Consequently, much discussion has concerned whether the GHQ-30 misses chronic illnesses where negative symptoms may be viewed as “usual” due to their longevity. Because the present study was conducted 6 months after stroke, patients were instructed to consider “usual” as their state of health before their strokes. The GHQ-30 was scored in the conventional 0-0-1-1 format, where any response indicating a deterioration from the usual is scored as 1. For the GHQ-30, the score taken is simply a total of these scores, giving a range of 0 to 30.

A self-completion form that included the HAD scale was left for return by mail. The HAD scale comprises 14 questions, of which half make up the anxiety subscale and half the depression subscale. The scale refers only to the patient’s feelings during the previous week and makes no reference to “usual” or past states. The response scale refers only to the patient’s feelings during the previous week, thus providing a severity and the level of severity in the previous week, thus providing an index of change. Supplementary questions were also administered to generate a DSM-IV diagnosis. Because it is a possible confounding variable in this physically ill sample, the fatigue rating scale was excluded. All indications from use in both the present and previous studies suggest that the SADS is both reliable and valid.

Two weeks later (mean, 14.2 days), a psychiatrist (S.M.) visited patients referred to our hospital with acute stroke were randomized. Of these, 16 died, 19 were severely cognitively impaired, and 7 refused to consent to follow-up, leaving 145 patients (77.5%) who were followed up by both the psychiatrist and psychologist at 6 months. The 145 subjects had a median age of 68 (range, 18 to 90 years), and 75 (51.7%) subjects were male. One hundred thirty-three patients (91.7%) completed the GHQ-30, and 111 (76.6%) the HAD Scale. Data were complete for both measures in 105 patients (72.4%). The primary causes of incomplete responses were inability to comprehend questions, refusal to answer specific questions, and failure of patients to return the self-completion form containing the HAD Scale (42% of those incomplete) or missing sections by turning over two pages at once. To estimate the size of any “nonresponse” bias introduced, we compared the baseline data of those in whom data were complete (n=105) with the remainder of those randomized (n=82). Patients in whom complete data were not collected were significantly more likely to have been dependent before the stroke, having suffered a severe stroke with cortical damage and cognitive impairment.

The SADS psychiatric evaluations of the 105 patients in whom data were complete identified 30 patients (28.6%) with 40 psychiatric diagnoses (Table; 7 patients had 2 diagnoses, and one had 4). The psychiatric evaluation of the 40 patients who failed to complete the study measures revealed that they were rather more likely to have a psychiatric diagnosis. There were 14 patients (35%) with 19 psychiatric diagnoses: 11 patients (27.5%) had depressive disorders, 3 patients (7.5%) had anxiety, and 5 patients (12.5%) had a variety of other disorders.

We compared the GHQ-30 and HAD Scale using ROC curves (Figure 1). No significant difference was found between the GHQ-30 and the HAD Scale total score to identify any DSM-IV case (z=-0.07, P=0.95, Figure 1). Neither was there any significant difference between the ability of the GHQ-30 and the HAD depression (z=-0.587, P=0.56) and anxiety (z=-1.155, P=0.25) subscales to detect cases of DSM-IV depression or anxiety, respectively.

The sensitivity and specificity rates for all cutoff points and grouped diagnoses for the GHQ-30 are illustrated in Figure 1. The recommended cutoff point, derived from a general practitioner sample, for the GHQ-30 is 4/5. Using this cutoff point in the present sample of stroke patients to identify all diagnoses produces a sensitivity of 0.9 and a specificity of 0.47. In this study, to gain a sensitivity of 0.5, on which the recommended cutoff point was based, a cutoff of either 13/14 or 14/15 would be necessary where the sensitivity is 0.53 and 0.47, respectively, and specificity is 0.89 and 0.91, respectively. The ROC curve suggests that for both a high sensitivity and specificity the best cutoff point is 8/9 in the present population, with a sensitivity of 0.8 and specificity of 0.76.

Selected Abbreviations and Acronyms

<table>
<thead>
<tr>
<th>Acronym</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>DSM-IV</td>
<td>Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition</td>
</tr>
<tr>
<td>GHQ</td>
<td>General Health Questionnaire</td>
</tr>
<tr>
<td>HAD Scale</td>
<td>Hospital Anxiety and Depression Scale</td>
</tr>
<tr>
<td>ROC</td>
<td>receiver operating characteristic</td>
</tr>
<tr>
<td>SADS</td>
<td>Schedule for Affective Disorders and Schizophrenia</td>
</tr>
</tbody>
</table>
Grouped Patient Diagnoses Defined by SADS With Supplementary Questions to Generate a Current DSM-IV Diagnosis

<table>
<thead>
<tr>
<th>Diagnosis</th>
<th>n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Any depressive disorder</td>
<td>20 (19)</td>
</tr>
<tr>
<td>Major depressive disorder</td>
<td>14 (13)</td>
</tr>
<tr>
<td>Depressive disorder not otherwise specified</td>
<td>1 (1)</td>
</tr>
<tr>
<td>Adjustment disorder with depressed mood</td>
<td>4 (4)</td>
</tr>
<tr>
<td>Dysthmic disorder</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Adjustment disorder with mixed anxiety and depressed mood*</td>
<td>1 (1)</td>
</tr>
<tr>
<td>Any anxiety disorder</td>
<td>13 (12)</td>
</tr>
<tr>
<td>Generalized anxiety disorder</td>
<td>2 (2)</td>
</tr>
<tr>
<td>Agoraphobia with or without panic disorder</td>
<td>8 (7)</td>
</tr>
<tr>
<td>Adjustment disorder with anxious mood</td>
<td>1 (1)</td>
</tr>
<tr>
<td>Adjustment disorder with mixed anxiety and depressed mood*</td>
<td>1 (1)</td>
</tr>
<tr>
<td>Specific phobia</td>
<td>1 (1)</td>
</tr>
<tr>
<td>Other psychiatric diagnoses</td>
<td>8 (7)</td>
</tr>
<tr>
<td>Alcohol dependence</td>
<td>6 (6)</td>
</tr>
<tr>
<td>Personality disorder</td>
<td>1 (1)</td>
</tr>
<tr>
<td>Sedative, hypnotic, or anxiolytic dependence</td>
<td>1 (1)</td>
</tr>
<tr>
<td>Total number of diagnoses</td>
<td>40</td>
</tr>
</tbody>
</table>

n indicates number of patients with a given diagnosis. Note: Of the 30 patients (with a total of 40 diagnoses), 29 had either a depressive or anxiety disorder. One patient had purely a personality disorder.

*One patient had mixed anxiety and depressed mood and was counted under both the “any anxiety disorder” and “any depressive disorder” categories.

The authors of the HAD scale unusually recommend two cutoff points, 8/9 for a high sensitivity and 10/11 for high specificity, for both their anxiety and depression subscales, allowing the practitioner to choose whether to include borderline cases.\textsuperscript{15} Using the 8/9 cutoff point in our patients for the depression subscale, identifying depression only, produced a sensitivity of 0.45 and a specificity of 0.85. A cutoff point of 10/11 produced a sensitivity of 0.35 and a specificity of 0.93. Improved sensitivity and specificity were achieved in this sample using a cutoff point of 6/7 (sensitivity, 0.8; specificity, 0.79).

For the HAD anxiety subscale (identifying anxiety cases only), a cutoff point of 8/9 produced a sensitivity of 0.5 and specificity of 0.87. A cutoff of 10/11 produced a sensitivity of 0.42 and specificity of 0.92. Again, as in the depression subscale, a better balance between sensitivity and specificity was achieved using a cutoff point of 6/7 (sensitivity, 0.83; specificity, 0.68). Figures for the summed scale are included in the present study only to facilitate comparison with previous studies (eg, Reference 20, Figure 1), although totaling of the scales was not recommended by the original authors.\textsuperscript{15}

To further facilitate comparison and choice of cutoffs periods, we calculated various cost ratios. Cost refers to the relative importance in different situations of a measure possessing either high sensitivity (ie, very few false-negatives) or high specificity (ie, very few false-positives). For example, in some situations it may be deemed far worse to miss a potentially treatable patient by using a measure with a low sensitivity than it would be to further examine a patient who is actually well by using a measure with a low specificity. The costs of each cutoff point have been calculated through a range of a false-negative (a patient missed), costing from one quarter to four times the cost of a false-positive (a well patient assessed for further treatment). That is the cost of a false-negative divided by the cost of a false-positive. For example, it may be considered twice as costly to miss a depressed patient than to assess a well patient for further treatment, corresponding to a ratio of two. Note that the estimated error cost depends not only on the sensitivity, specificity, and costs of false-negative and false-positive errors but also on the prevalence of the condition in the population. Explicitly,

\[
\text{Total Cost} = \pi (1-\text{sensitivity})C_{FN} + (1-\pi)(1-\text{specificity})C_{FP}
\]

where \(\pi\) is the prevalence and \(C_{FN}\) and \(C_{FP}\) are the costs of false-negative and false-positive errors, respectively. This implies that simply choosing a threshold that “balances” both sensitivity and specificity in some way does not imply an assumption of equal error costs. Based on this formula, the estimated optimal cutoff points for various cost ratios are plotted in Figures 2 and 3 for the GHQ and HAD, respectively.

**Discussion**

It is important for both clinicians and researchers to reliably identify mood disorders after stroke. Poststroke depression is a common and debilitating disorder that may slow rehabilitation and produce a permanent negative influence on recov-
Early screening and identification of mood disorders may be important if an effective treatment exists. In addition, large randomized controlled trials of treatment that aim to influence psychological outcomes require reliable self-report measures in which knowledge of both sensitivity and specificity is necessary to compute the power of the study and to facilitate the choice of cutoff point.

This study identified a fairly representative sample of hospital-referred stroke patients similar on all indices measured to all the patients assessed at our hospital during the study period. The necessity for patients to be hospital referred may have resulted in extremely mild and severe strokes being underrepresented. Patients who suffered severe cognitive impairment or who were unable to communicate effectively were excluded; while we would acknowledge that due to these impairments they might be at greater risk of depression, self-report measures are clearly an inappropriate method of assessment for this group. Furthermore, strokes that did not merit hospital referral might have a correspondingly low frequency of mood disorders. Thus, our sample may represent a "middle ground" of stroke severity, but this is precisely the population in whom such measures would be most appropriate in clinical and research practice. Also, because the frequency of mood disorder varies over the months following a stroke, it would be unwise to generalize our findings to situations in which the frequency may be very different (eg, soon after stroke).

Some might criticize our choice of the GHQ-30, which was designed to measure the overall burden of psychiatric symptoms rather than (as is the case with the 28-item version) to identify patients likely to have specific psychiatric diagnoses. However, on balance we opted for the 30-item version since it was designed to be used in physically ill people whose somatic symptoms might reduce the validity of other versions. Even this version, however, refers to "sleep," "chatting," and "getting out," which might reflect physical as well as psychiatric problems. This could in part account for the increased rates of positive response in our population in comparison with the general practitioner sample previously used for validation.

Other methodological factors potentially confound our comparison of the two measures. First, our two measures were delivered in different ways. The GHQ-30 was administered by a research associate who read out each question and recorded patients’ answers for them. The GHQ-30 is normally completed by patients independently but was not done so in this case to achieve uniformity with the remainder of the structured interview. The HAD Scale was left with patients for self-completion. This was reflected in the substantially higher completion rate for the GHQ-30 (92%) compared with...
the HAD Scale (77%). However, given the different methods of delivery, our data cannot support the hypothesis that the GHQ-30 is a more practical measure or one that would be associated with a higher completion rate than the HAD Scale. Indeed, we suspect that given its complexity and greater length, the response rates to a self-completed GHQ-30 would be no better and possibly worse than for the HAD Scale. Particularly relevant for a population 6 months after stroke is the criticism that the GHQ-30 misses chronic cases because of its reference to a “usual” state.27 We hoped that our instructions to regard “usual” as health status before stroke would partially overcome this, but found that patients had difficulty remembering pre-stroke health. Some patients, and particularly those who are depressed, might have a rather negative view of their pre-stroke status, whereas others might have an overly positive view of pre-stroke status. This could distort our results in either direction and in an unpredictable way.

The two scales were often completed a few days apart, which may confound our comparisons if there were systematic changes in the patients’ psychiatric state over these few days. Significant bias seems unlikely, especially since these patients were assessed at least 6 months after their stroke, when we considered mood likely to have stabilized (a view supported by the psychiatric assessments, in which no patients reported changes in their mental state within the preceding fortnight). Ideally, the measures would have been completed at the same time, but we judged that patients would be neither able nor willing to complete three psychiatric measures at one time.

One previous comparison28 of the GHQ and the HAD Scale in stroke reported that the 28-item version of the GHQ (n=66) was superior to the HAD Scale (n=93) in detecting both anxiety and depression. Similar studies have been conducted in other medically ill populations. Lewis and Wessely20 found no difference between the GHQ-12 and the summed HAD Scale in detecting cases of minor psychiatric disorder in a sample of dermatological patients. Wilkinson and Barczak,29 in a general practitioner sample, found that the HAD Scale was generally more sensitive and simpler to complete than the GHQ-28. Aylard et al30 undertook a further validation of both the HAD Scale and the anxiety and depression subscales of the GHQ-28 in a hospital outpatient sample. They found both to be suitable for preliminary screening, and suggested the use of a borderline range, a score range where patients are “bordering” on being considered a “case.” in the GHQ.

When considering which measure should be recommended for what purpose, it is useful to consider the balance of sensitivity and specificity at different cutoff points (eg, Figure 1). However, because there was little difference between the two measures, any choice may have to be based on the measures’ practicality, acceptability to patients, and whether similar studies have used one or other, which might facilitate future systematic review.

The recommended cutoff points for the GHQ-30 and HAD Scale appear suboptimal in our group of stroke patients. When one considers which cutoff is most appropriate for a given population or use, the comparative cost of a false-positive or false-negative in those circumstances might usefully be taken into account. For example, in a clinical setting where it is most undesirable to miss treatable cases and psychiatric resources are not too limited, a false-negative may be deemed to cost four times more than a false-positive. Reference to Figures 2 and 3 illustrates that at point 4 on the horizontal axis, the optimal cutoff point for identifying depression is 9/10 for the GHQ and 6/7 for the HAD. We would suggest that to facilitate a decision regarding cutoff points, potential users consider the comparative costs within their frame of use and choose the optimum cutoff for their cost ratio as specified in Figures 2 and 3.

In conclusion, the GHQ-30 and HAD Scale appeared to differ little in terms of their sensitivity and specificity for diagnosing mood disorders 6 months after stroke, although the HAD Scale was significantly shorter and, we suspect, may have been easier for patients to complete. Recommended cutoff points may not offer the best balance between sensitivity and specificity in this group of patients. We hope that our data, and especially those referring to cost ratios, will be useful to others who plan to use these measures to screen their patients for psychiatric problems or who wish to use them as measures of outcome in randomized trials.

Acknowledgments

We acknowledge the generous support and funding received from the Scottish Home and Health Department, Stroke Association, and Medical Research Council.

References

Detecting Psychiatric Morbidity After Stroke: Comparison of the GHQ and the HAD Scale
Suzanne O'Rourke, Siobhan MacHale, David Signorini and Martin Dennis

Stroke. 1998;29:980-985
doi: 10.1161/01.STR.29.5.980
Stroke is published by the American Heart Association, 7272 Greenville Avenue, Dallas, TX 75231
Copyright © 1998 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/29/5/980

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/