

# Validation of an outcome scale for use in adult psychiatric practice

Roy McClelland, Peter Trimble, Michael L Fox, Michael R Stevenson, Beverley Bell

## Abstract

**Objective**—To clarify the usefulness, acceptability, sensitivity, and validity of version 4 of the Health of the Nation Outcome Scale (HoNOS), a scale developed to meet the requirement for a clinically acceptable outcome scale for routine use in mental illness services.

**Design**—Patients with a range of mental illnesses were rated on the HoNOS at the beginning and end of an episode by interviewers with mental health professionals.

**Subjects**—934 patients from eight diagnostic categories were rated by 129 mental health professionals at 17 sites; 250 were also rated on a range of comparison scales.

**Outcome measures**—Comparison of patients' scores at the beginning and end of an episode using individual item scores, dimensional subscores, and the total score.

**Results**—HoNOS scores decreased by almost 50% between the beginning and end of episodes. They varied with the severity of the setting and discriminant analysis showed that the HoNOS had a moderate level of discriminatory power. Correlation analysis showed acceptable levels of agreement with independent scales, although the accuracy of ratings of some items at the beginning of an episode was affected by information deficits.

**Conclusion**—The findings indicate that HoNOS is sensitive to change across time and to differences in illness type and severity, and has a sufficient degree of both construct and criterion related validity to fulfil the requirements of a mental health outcome scale for routine use in clinical settings.

(Quality in Health Care 2000;9:98–105)

**Keywords:** Health of the Nation Outcome Scale (HoNOS); mental illness; validation; outcome measures

Interest in the quality and effectiveness of health services has grown considerably over the past decade. In this regard UK health departments, health funding agencies, government, and other bodies involved in the commissioning of care have been concerned with the lack of information on health outcomes. Information regarding the outcome of various interventions in health care has largely been based on efficacy research from clinical trials carried out under controlled research conditions. What is required is more and better information related to the effectiveness of treatments in

routine clinical practice.<sup>1</sup> Clinical outcome embraces a number of dimensions including symptomatology, social functioning, patient satisfaction, well being, and health related quality of life. In relation to mental health practice, our review of the literature revealed significant shortcomings in most available scales. General measures such as the Short Form 36<sup>2</sup> would not provide adequate detail regarding specific conditions in mental health. Other measures designed to assess levels of dysfunction in specific mental illness groups appeared to have less value when applied to broader groups because of likely ceiling or floor effects. To fill this gap the UK Royal College of Psychiatrists' Research Unit, the professional and educational body for psychiatrists in the UK and Ireland, developed a new scale—the Health of The Nation Outcome Scale (HoNOS). The development of the scale and a preliminary assessment of the reliability and criterion validity of the most recent version, HoNOS version 4 (HoNOS-Severe Mental Illness (SMI)) have been described by Wing and colleagues.<sup>3,4</sup>

The HoNOS-SMI is a 12 item scale designed to provide a brief, accurate, and relevant measure of mental health and social functioning (box 1). Each item measures a type of problem commonly presented by patients in mental healthcare settings and each is scored on a five point scale ranging from 0 (no problem) to 4 (severe/very severe problem). The 12 items are intended to cover four areas of mental health: behaviour (1–3), impairment (4 and 5), symptoms (6–8), and social functioning/context (9–12). Patients are rated on the HoNOS when sufficient information has become available. Ratings are carried out either by a single practitioner or using input from the clinical team. Outcome is measured by comparing a patient's scores at two points in time,

- |   |
|---|
| <ol style="list-style-type: none"> <li>1 Overactive, aggressive, disruptive</li> <li>2 Non-accidental self-injury</li> <li>3 Problem drinking or drug taking</li> <li>4 Cognitive problems</li> <li>5 Physical illness or disability problems</li> <li>6 Problems with hallucinations and delusions</li> <li>7 Problems with depressed mood</li> <li>8 Other mental and behavioural problems (specify problem)</li> <li>9 Problems with relationships</li> <li>10 Problems with activities of daily living</li> <li>11 Problems with living conditions</li> <li>12 Problems with occupation and activities</li> </ol> |
|---|

Box 1 Health of the Nation Outcome Scale: score sheet

Department of Mental Health, The Queen's University of Belfast, Belfast BT7 1NN, UK  
 R McClelland, professor of mental health  
 P Trimble, research fellow  
 M L Fox, research assistant  
 B Bell, research nurse

Health and Social Care Research Unit, The Queen's University of Belfast, Belfast BT7 1NN, UK  
 M R Stevenson

Correspondence to:  
 Professor R McClelland  
 Email:  
 R.J.McClelland@qub.ac.uk

Accepted 17 January 2000

using individual item scores, the dimensional subscores, and the total score.

The present study extended the work of the College to an evaluation of the properties, sensitivity, and construct validity of the HoNOS-SMI and to a more comprehensive exploration of the criterion validity. Qualitative feedback from HoNOS raters was obtained through a series of interviews. Particular emphasis was placed on assessing the functioning of the HoNOS-SMI under conditions representative of routine use within mental illness services. This work was commissioned by the Department of Health and Social Services (NI) to fulfil one of the targets of the NI Regional Strategy for Mental Health, the development of a clinically acceptable outcome scale for routine use in mental illness services.

## Methods

### RATERS AND TRAINING

Raters were drawn from all mental health professionals and included psychiatrists, trainee psychiatrists, psychiatric nurses, occupational therapists, and social workers. Most raters were either doctors or nurses, reflecting clinical practice situations including availability of staff. Other professions contributed to multidisciplinary ratings. The time to complete the questionnaire was 5–15 minutes, depending on the experience of the rater and the complexity of the patient's problems.

Two members of the project team were trained in the use of the HoNOS at the Research Unit of the Royal College of Psychiatrists. Key staff from participating sites were trained by the Project Team in a series of one day workshops. Each workshop included training in the use of the HoNOS and instruction on the rationale and usefulness of outcome measurement. The Project Team visited hospital and community groups on request to provide guidance during the introduction of the HoNOS.

### COLLECTION OF DATA

Patient ratings were obtained from a range of treatment environments including community and outpatient services, day hospitals, acute wards, and intensive care units. Staff at selected sites were asked to complete an HoNOS rating for each patient as part of routine assessment at the beginning of a spell of illness and, where feasible, at the end. This approach allowed a gradual introduction of the instrument, provided a means of gauging the acceptability of the scale to the staff concerned, and enabled the Project Team to gather a large amount of initial data (time 1) on which to investigate construct validity. Second ratings (time 2) were to be made by the same raters and it was recognised from the outset that staff changes, availability, and patient access would reduce the number of time 2 observations.

Ratings were made according to guidelines produced by the Royal College of Psychiatrists.<sup>5</sup> An initial rating was made at the beginning of the treatment episode based on the two weeks prior to the assessment. Ratings were based either on a team discussion, provid-

ing a consensus rating, or on assessment by an individual professional. A second rating was made when clinically indicated—for example, at the end of a treatment episode, before discharge from a programme of care or unit, or following a significant clinical development. For patients in a continuing care setting second ratings were completed approximately three months after the first rating.

For each patient the HoNOS rater completed a background information sheet. This included the identity and profession of the rater, the setting where the rating took place, and whether the rating was carried out by an individual or by a team. It also detailed relevant information about the subject, including diagnostic category, Mental Health Order status, date of birth, and marital and employment status.

### COLLECTION OF COMPARISON SCALE DATA

To determine the criterion validity of the HoNOS, a number of comparison scales were selected to provide comprehensive coverage of mental illness symptoms and social functioning. This was carried out on a sample of patients rated by staff on the HoNOS. The aim was to achieve a large enough sample across a range of diagnostic groups on which a comparative analysis could be made; 250–300 patients were considered sufficient. All patients providing comparison data were rated on the Global Assessment Scale (GAS),<sup>6</sup> the Social Role Performance Schedule (SRPS)<sup>7</sup> and, initially, on one of three psychiatric questionnaires—namely, the Brief Psychiatric Rating Scale 18 (BPRS),<sup>8</sup> the Hamilton Rating Scale for Depression (Ham-D),<sup>9</sup> or the Hamilton Anxiety Scale (Ham-A).<sup>10</sup> The use of the questionnaires was based on diagnosis. Four broad diagnostic groups were employed for this purpose: Psychosis, Depression, Neurosis, and Other. Allocation of subjects to each category was made by clinical teams or individual HoNOS raters. Through a process of weighted randomisation most of the Psychosis group were rated on the BPRS, the Depression group on the Ham-D, and the Neurosis group on the Ham-A. For the Other group equal weight was given to the two Hamilton scales. Because the return of data was slower than anticipated and the time allocated to the project limited, the use of the Hamilton scales was discontinued and greater emphasis was placed on the BPRS in conjunction with the GAS and the SRPS.

Patients were rated on the comparison scales within three days of the HoNOS assessment. Having obtained consent, semi-structured interviews were conducted on site by raters trained in the use of these instruments.

### SUBJECTS

Patients who took part in the study were selected to provide a range of mental health problems of sufficient severity to warrant referral for treatment by specialist mental health services, thus reflecting the population for which the HoNOS-SMI was intended. A total of 934 subjects participated in the project (368 men) with ages ranging from 16 to 64 years. Of

Table 1 Characteristics of study subjects at time 1 and time 2

Characteristic	Time 1		Time 2	
	n	%	n	%
Diagnosis				
Dementia	2	0.2%	0	0.0%
Drug/alcohol	51	5.5%	36	10.4%
Psychosis	207	22.2%	101	29.2%
Depression	276	29.6%	112	32.4%
Manic/bipolar	77	8.2%	37	10.7%
Neurosis	93	10.0%	26	7.5%
Eating/stress	100	10.7%	8	2.3%
Personality	83	8.9%	17	4.9%
Other	26	2.8%	9	2.6%
Mental health status				
Detained	142	15.2%	84	24.3%
Voluntary	778	83.3%	258	74.6%
Setting				
ICU	74	8.0%	40	11.5%
Acute	398	42.8%	245	70.8%
Day hospital	151	16.1%	53	15.3%
OP/community	309	32.3%	6	1.8%

the total sample 346 were rated a second time (176 men). The characteristics of the sample at time 1 and time 2 are shown in table 1.

#### STATISTICAL ANALYSES

Spearman rank correlation coefficients were calculated to examine the HoNOS items for unnecessary duplication of function. To examine the internal consistency of the HoNOS the 12 items were intercorrelated. As the HoNOS was constructed with 12 ordered category scales, the most appropriate measure of association between individual HoNOS items was the non-parametric Spearman rank method. The dimensionality of the HoNOS was explored using principal components analysis and a range of mean scores was calculated to establish evidence of both the construct validity and the sensitivity of the instrument. Further evidence of construct validity was provided by discriminant analyses. In order to establish the criterion validity Spearman rank correlations were used to determine the relationships between the HoNOS and the comparison scales. As there were a number of factors that were likely to have a negative effect on the associations between the scales—that is, the large number of HoNOS raters providing data, differences in the way data were collected from the HoNOS and comparison scales, and the fact that the instruments were not wholly equivalent—correlation coefficients of 0.6 and above were taken to indicate satisfactory performance on the part of the outcome scale. The Student's *t* test was used to assess the significance of a change in scores between times 1 and 2. Small variations in the total number of patients for some calculations were the result of missing data.

#### Results

A total of 129 raters from 17 sites completed HoNOS ratings in the course of the project. The numbers of HoNOS ratings at times 1 and 2 for each rater occupation are summarised in table 2. At time 1, 439 HoNOS ratings (48%) were known to have been completed on the basis of a team consensus and 475 ratings (52%) by raters operating on an individual

Table 2 Patients rated by rater occupation

Rater occupation	Patients rated		Percentage	
	Time 1	Time 2	Time 1	Time 2
Nurse	657	293	70.3%	84.7%
CPN	50	2	5.4%	0.6%
Doctor	139	34	14.9%	9.8%
Occupational therapist	8	2	0.9%	0.6%
Social worker	43	1	4.6%	0.3%
Other	37	13	4.0%	3.8%

CPN = community psychiatric nurse.

basis. At time 2, 133 team ratings (40%) and 197 single ratings (60%) were completed. Comparison scales were completed on 250 subjects. All were rated on the GAS and the SPS, 133 were rated on the PRS, 85 on the Ham-D, and 33 on the Ham-A. Of these subjects 138 were rated a second time within three days of being rated on HoNOS at time 2. All of these were rated on the GAS and the SPS, 83 were rated on the BPRS, 42 on the Ham-D, and 14 on the Ham-A. Loss of time 2 comparisons resulted from patients being discharged from care at relatively short notice or staff not being available for second HoNOS ratings.

#### PROPERTIES OF HONOS-SMI

Each of the 12 HoNOS items is intended to quantify discrete facets of mental health without unnecessary duplication. This feature was examined using Spearman rank correlations. The resulting coefficients were generally low and positive with only two exceeding 0.4. This suggested that duplication of function among the HoNOS items was small and that the instrument had very little redundancy among scale items.

Principal components analysis was applied to the HoNOS time 1 data to explore the dimensionality of the scale. Four discrete factors were identified, accounting for 55% of the total variance. The first factor ("severity of illness") encompassed eight of the 12 items including the four social items. HoNOS item 2 (non-accidental self-injury) and item 7 (depressed mood) were extracted together as a second factor, reflecting the close association between these two aspects of mental health. Item 5 (physical illness/disability) was extracted on its own, thus differentiating physical problems from mental health problems. Item 8 (other mental and behavioural problems) was also extracted as a distinct factor, reflecting the breadth of the conceptual domain it encompasses. This item incorporates nine separate symptoms including anxiety, stress, eating disorder, sleep disorder, and sexual dysfunction.

To examine the weight of item scoring the percentage contributions of each item to the overall mean HoNOS total were calculated (table 3).

Given that the scale uses 12 items, each item would, in statistical terms, be expected to contribute approximately one twelfth or 8.3% to the total score. It was found that item 7 (depressed mood), item 8 (other mental and behavioural problems), and item 9 (problems with relationships) were scored high relative to

Table 3 Item contribution to HoNOS total score

Item no.	Item title	Contribution to total
1	Aggression/overactivity	8.4%
2	Self-harm	7.4%
3	Substance abuse	7.3%
4	Cognitive problems	4.8%
5	Physical illness/disability	4.8%
6	Hallucination/delusion	7.3%
7	Depressed mood	14.6%
8	Other mental/behavioural problems	18.7%
9	Relationships	14.2%
10	Activities of daily living	6.7%
11	Living conditions	3.3%
12	Occupation and activities	3.4%

the other items, contributing 14.6%, 18.7%, and 14.2%, respectively, to the total. This was not unexpected given that these items quantify problems that are common to most psychiatric disorders. The magnitude of the contribution of item 8 reflects the broad spectrum of symptoms included. By contrast, item 11 (living conditions) and item 12 (occupation and activities) contributed only 3.3% and 3.4%, respectively, to the mean HoNOS total, indicating that patients were generally scored low on these items.

#### CONSTRUCT VALIDITY AND SENSITIVITY

The construct validity and sensitivity of HoNOS-4 was assessed by examination of the variation in HoNOS scores across time and with setting, mental health status, and diagnostic groups. Discriminant analysis was used for further exploration.

The mean HoNOS total score decreased by 49% from time 1 to time 2 (table 4). All changes for the HoNOS subtotals and total were significant ( $p < 0.001$ ). With respect to the subtotal scores, the largest absolute change occurred on the symptom items, followed by behaviour, social functioning, and impairment. The mean scores for all of the individual HoNOS items were reduced significantly from time 1 to time 2 ( $p < 0.001$ ), with the exception of item 11 ( $p > 0.05$ , NS). As indicated by the examination of item weights, the highest mean item score at time 1 was recorded for item 8, followed by items 7 and 9. The lowest scores were recorded for items 12, 5, and 11.

With respect to settings, the mean HoNOS total scores at time 1 decreased as the level of

illness severity anticipated within these settings decreased (table 5). The highest mean total was obtained for intensive care units followed by acute wards, day hospitals, and outpatient/community settings. The same pattern of a stepwise decrease occurred for eight of the 12 items, the exceptions being items 5, 7, 8 and 9. Further computation showed that the greatest change between times 1 and 2 occurred with intensive care units ( $\bar{x} = -9.85$ ) followed by acute wards ( $\bar{x} = -6.56$ ) and day hospitals ( $\bar{x} = -2.00$ ). Change within outpatient/community settings could not be calculated because of a low return of time 2 data.

Patients detained under the Mental Health Order had significantly higher mean HoNOS total and subtotal scores at time 1 than those of voluntary patients ( $p < 0.001$ ), reflecting assumed greater illness severity (table 6). At item level, detained patients scored higher on nine of the 12 items, particularly on item 1 (aggression/overactivity) and item 6 (hallucinations/delusions). Detained patients had a greater mean change in scores between times 1 and 2, reflecting their higher scores at time 1.

When HoNOS time 1 scores were calculated separately for each diagnostic group the highest mean total score was obtained for the drug/alcohol group, reflecting wide dysfunction across Behavioural, Symptom and Social Functioning domains. High scores were also obtained for Psychosis and Bipolar Disorder groups, followed by Depression, Personality Disorder, Neurosis and Eat/Sleep/Stress disorders (table 7).

Each of the groups scored most highly on the HoNOS item most relevant to their disorder, indicating that the scale has some capacity to discriminate between diagnostic groupings. This facet of the scale was explored further using discriminant analysis which was employed to determine whether a patient's diagnostic grouping could be predicted on the basis of his/her profile of HoNOS item scores. With the application of discriminant functions, 57% of the cases were correctly predicted, equating to a kappa ( $\kappa$ ) value of 0.47, thus demonstrating that HoNOS has a moderate level of discriminatory power (table 8). Four HoNOS items (4, 5, 11, and 12) were found to have no discriminatory power and thus played no part in identifying different diagnostic groups. It was found that successful prediction was highest among patients in the Psychosis, Depression, and the Eating/Sleep/Stress groups and lowest among the Manic/Bipolar, Neurosis/Anxiety, and Personality groups. Removing the Manic/Bipolar and Neurosis/Anxiety cases from the analysis increased the level of successful prediction to 70% ( $\kappa = 0.58$ ).

#### CRITERION RELATED VALIDITY

HoNOS associations with the BPRS and the GAS were evaluated using Spearman rank correlations. The relationships between the scales at times 1 and 2 were examined separately and differences in the performance of HoNOS were found. The time 1 correlations reached a mod-

Table 4 Mean HoNOS scores at times 1 and 2

HoNOS items and totals	n	Time 1	Time 2	Change
H1 Aggression/overactivity	346	1.15	0.49	-0.66 (58%)
H2 Self-harm	346	1.01	0.32	-0.70 (69%)
H3 Substance abuse	346	0.95	0.31	-0.65 (68%)
H4 Cognitive	345	0.73	0.37	-0.36 (49%)
H5 Disability	346	0.50	0.38	-0.12 (25%)
H6 Hallucinations/delusions	346	1.22	0.53	-0.69 (57%)
H7 Depression	346	1.65	0.72	-0.92 (56%)
H8 Other mental/behavioural problems	344	1.93	0.97	-0.97 (50%)
H9 Relationships	344	1.58	1.11	-0.47 (30%)
H10 Daily living	346	1.02	0.58	-0.44 (43%)
H11 Living conditions	335	0.51	0.43	-0.08 (15%)
H12 Occupation/activities	335	0.49	0.27	-0.22 (45%)
H1-3 Behaviour total	346	3.11	1.11	-2.00 (64%)
H4-5 Impairment total	346	1.23	0.75	-0.48 (39%)
H6-8 Symptom total	346	4.76	2.21	-2.55 (54%)
H9-12 Social total	346	3.57	2.36	-1.21 (34%)
H1-12 HoNOS total	346	12.67	6.43	-6.24 (49%)

Table 5 Mean HoNOS item scores at time 1 by setting

HoNOS items and totals		Intensive care (n=74)	Acute wards (n=398)	Day hospital (n=151)	OP/ community (n=302)
H1	Aggression/overactivity	2.39	0.90	0.79	0.58
H2	Self-harm	1.41	0.97	0.82	0.42
H3	Substance abuse	1.53	0.88	0.60	0.58
H4	Cognitive	1.58	0.53	0.48	0.26
H5	Disability	0.73	0.46	0.52	0.54
H6	Hallucinations/delusions	1.69	1.11	0.67	0.17
H7	Depression	1.43	1.64	1.93	1.35
H8	Other mental and behavioural problems	2.00	1.79	2.09	2.31
H9	Relationships	1.91	1.48	1.72	1.40
H10	Daily living	1.50	0.93	0.55	0.31
H11	Living conditions	0.85	0.47	0.23	0.13
H12	Occupation/activities	1.26	0.44	0.20	0.17
H1-3	Behaviour total	5.32	2.75	2.21	1.58
H4-5	Impairment total	2.26	0.99	1.00	0.80
H6-8	Symptom total	4.95	4.45	4.66	3.75
H9-12	Social total	5.20	3.30	2.87	2.00
H1-12	HoNOS total	17.73	11.56	10.56	8.13

Table 6 Mean HoNOS time 1 and change scores by mental health status

HoNOS items and totals		Time 1		Change scores	
		Voluntary (n=778)	Detained (n=142)	Voluntary (n=258)	Detained (n=84)
H1	Aggression/overactivity	0.73	1.93	-0.40 (55%)	-1.46 (76%)
H2	Self-harm	0.76	1.02	-0.66 (87%)	-0.85 (83%)
H3	Substance abuse	0.74	1.13	-0.58 (78%)	-0.89 (79%)
H4	Cognitive	0.43	1.02	-0.23 (55%)	-0.73 (71%)
H5	Disability	0.53	0.48	-0.10 (18%)	-0.19 (41%)
H6	Hallucinations/delusions	0.57	1.99	-0.46 (80%)	-1.40 (70%)
H7	Depression	1.64	1.25	-0.95 (58%)	-0.87 (69%)
H8	Other mental and behavioural problems	2.05	1.87	-0.93 (46%)	-1.03 (55%)
H9	Relationships	1.47	1.81	-0.39 (27%)	-0.74 (41%)
H10	Daily living	0.62	1.26	-0.36 (59%)	-0.67 (53%)
H11	Living conditions	0.30	0.66	-0.05 (18%)	-0.21 (31%)
H12	Occupation/activities	0.31	0.65	-0.20 (64%)	-0.31 (48%)
H1-3	Behaviour total	2.22	4.08	-1.63 (73%)	-3.20 (78%)
H4-5	Impairment total	0.96	1.49	-0.33 (35%)	-0.92 (62%)
H6-8	Symptom total	4.21	5.03	-2.33 (55%)	-3.23 (64%)
H9-12	Social total	2.69	4.30	-0.99 (37%)	-1.90 (44%)
H1-12	HoNOS total	10.08	14.90	-5.28 (52%)	-9.25 (62%)

est level,  $r_s = 0.44$  ( $p < 0.001$ ) for the BPRS and  $r_s = 0.49$  ( $p < 0.001$ ) for the GAS, while those at time 2 were high,  $r_s = 0.72$  ( $p < 0.001$ ) for the BPRS and  $r_s = 0.71$  ( $p < 0.001$ ) for the GAS (Table 9). When the changes in HoNOS scores from time 1 to time 2 were correlated with the changes in the comparison scales the coefficients exceeded 0.6 ( $p < 0.001$ ), indicating that the HoNOS has comparable dynamic properties and a similar capacity to record change.

The effects of a number of variables on the relationships between the scales were explored, including diagnosis, setting, and team versus single raters. As shown in table 9, the time 2 correlations for these variables were larger than those at time 1, with the exception of those based on ratings obtained from day hospitals. In most cases the time 2 coefficients reached and exceeded the 0.6 criterion while all but one of the time 1 coefficients did not. The low day hospital correlation between the HoNOS and BPRS at time 2,  $r = 0.42$  ( $p > 0.05$ , NS), was accounted for by the small number of cases included in the calculation ( $n = 15$ ) and by a disparity between BPRS and HoNOS scores for two of these cases. Overall, these results suggested that variation with diagnostic groups, settings, and team/single ratings did not account for the lower correlations at time 1.

Further analysis indicated that data from one of the acute wards impinged negatively on the overall time 1 HoNOS/BPRS correlation (table 9). When data from this ward were omitted from the calculation the coefficient increased from  $r_s = 0.44$  ( $p < 0.001$ ) to  $r_s = 0.56$  ( $p < 0.001$ ) and, when excluded from the all cases calculation, it increased from  $r_s = 0.59$  ( $p < 0.001$ ) to  $r_s = 0.63$  ( $p < 0.001$ ). This site differed from the others in two respects. Firstly, it had by far the largest number of raters ( $n = 16$ ) on a single site and, secondly, the key trained rater was absent for the greater part of the study and was thus unavailable to provide instruction and feedback to colleagues. It is likely that these factors contributed to an increased variation in the HoNOS ratings.

Exploration of diagnostic groups showed that HoNOS/BPRS coefficients for the Psychosis and Depression groups at time 1 exceeded 0.5, but the correlation for the other diagnostic groups combined ( $n = 43$ ) was substantially lower,  $r_s = 0.23$  ( $p > 0.05$ , NS). It is therefore clear that the data from the other diagnostic groups played some role in driving the overall time 1 correlation downward.

Further explorations were undertaken with respect to HoNOS items and their associations with individual items within the comparison scales. The resulting correlations largely reflected the degree of equivalence between items, with the coefficients at time 1 being lower than those at time 2. The four social items correlated poorly with their working equivalents in the SPS at both time 1 and time 2 (table 10). An explanation for this, and for the reduced time 1 correlations, was provided by HoNOS raters in the course of interviews (see below).

#### INTERVIEWS WITH HONOS RATERS

Interviews with HoNOS raters were conducted to assess the HoNOS in terms of its usefulness and acceptability to staff who used the instrument routinely as part of their day to day work and to highlight difficulties associated with the use of the scale.

Interviews took place at 10 of the sites participating in the project. The number of raters attending each interview ranged from one to five. The interviews were semi-structured and varied in length from 30 to 60 minutes. Raters were assured that their contributions to the discussions would remain confidential.

When asked about their overall assessment of the scale staff from eight of the sites responded positively and two responded negatively. The sites which regarded the HoNOS positively suggested that they would be supportive of any initiative to introduce the HoNOS more widely within mental illness services. The scale was described as "useful", "user friendly", "beneficial", and "good for quantifying illness and change". The HoNOS "validates care giving", it "promotes an objective view of patients", "highlighted patient problems quickly", "indicates the level of risk and improvement", "benefits the patient in the long run", "would be useful for monitoring

Table 7 Mean HoNOS item scores at time 1 by diagnosis

HoNOS items and totals		Drug/ alcohol (n=51)	Psychosis (n=207)	Depression (n=276)	Bipolar (n=77)	Neurosis (n=93)	Eat/stress (n=100)	Personality (n=83)
H1	Aggression/overactivity	1.47	1.16	0.69	1.42	0.63	0.35	1.28
H2	Self-harm	1.26	0.70	1.04	0.55	0.57	0.28	1.22
H3	Substance abuse	2.96	0.76	0.65	0.49	0.51	0.34	1.08
H4	Cognitive	0.51	0.86	0.47	0.94	0.29	0.11	0.29
H5	Disability	0.71	0.33	0.55	0.62	0.71	0.44	0.48
H6	Hallucinations/delusions	0.67	2.24	0.30	1.09	1.51	0.12	0.18
H7	Depression	1.35	1.26	2.20	1.17	1.73	0.97	1.45
H8	Other mental and behavioural problems	1.43	1.68	2.02	1.80	2.50	3.08	1.65
H9	Relationships	1.53	1.58	1.62	1.44	1.29	1.05	2.23
H10	Daily living	1.02	1.04	0.71	1.12	0.55	0.12	0.45
H11	Living conditions	0.75	0.61	0.27	0.26	0.24	0.10	0.24
H12	Occupation/activities	0.75	0.56	0.31	0.49	0.23	0.15	0.15
H1-3	Behaviour total	5.69	2.62	2.38	2.46	1.71	0.97	3.58
H4-5	Impairment total	1.22	1.19	1.02	1.56	1.00	0.55	0.76
H6-8	Symptom total	3.39	5.07	4.48	4.04	4.36	4.14	3.28
H9-12	Social total	3.94	3.75	2.88	3.30	2.30	1.42	3.06
H1-12	HoNOS total	14.24	12.63	10.76	11.35	9.37	7.08	10.67

Table 8 Discriminant analyses: predicted diagnoses based on HoNOS profiles

Primary diagnosis	Predicted cases							Actual cases
	Drug/ alcohol	Psychosis	Depression	Manic/ bipolar	Neurosis/ anxiety	Eat/sleep/ stress	Personality	
Drug/alcohol abuse	33	6	6				4	49
Psychosis	6	143	27	9		4	5	194
Depression	16	21	217	1	2	8	7	272
Manic/bipolar	3	21	29	10	1	9	3	76
Neurosis/anxiety	1	3	68	1	4	15		92
Eating/sleep/stress	1	4	21	1	1	66	3	97
Personality	9	3	42	5		4	19	82
Total	69	201	410	27	8	106	41	862

purposes”, and “would show that patients get better in our care”. However, a number of reservations were expressed at most sites, suggesting the need for some modifications to the scale or in its use.

The two dissenting sites suspected HoNOS was “a blunt instrument” that was “open to misinterpretation” and may not reflect a patient’s mental health. These issues were part of the present objective evaluation.

Item 8 was a particular source of dissatisfaction as only one of the nine symptoms subsumed by this item could be scored during rating. A common view was that important information about the patients was often omitted because of this constraint. The social items, particularly items 10, 11 and 12, were problematic because the information necessary to

score them effectively was often unavailable. It was strongly suggested that community personnel such as social workers or community psychiatric nurses would rate these items more appropriately. Nine of the 10 sites suggested that time 1 ratings were less accurate than time 2 ratings because more information tended to be disclosed by patients as rapport with staff improved over time.

At some of the sites raters were not provided with time to devote specifically to the rating of patients. Consequently, raters often completed ratings in their own time, which had a subsequent effect on their attitude and motivation, and possibly on the quality of the ratings.

In addition, the interviews revealed that guidelines for the use of the HoNOS were in some cases not adhered to. At the outset of the project raters were instructed to rate patients at the beginning and the end of an episode of care and that ratings were to be based on the preceding two weeks. Some deviation from these instructions was evident, which may indicate a need for closer supervision of raters. Part of this deviation may be accounted for by the heavy demands placed on staff by their day to day responsibilities and the absence of specific time allocated for HoNOS ratings.

These findings indicate that the HoNOS is likely to be regarded as a user friendly instrument if used routinely as an outcome measure for mental illness services. However, the experience of the raters suggested that several modifications to the scale, to the rating guidelines, and to on-site staff support arrange-

Table 9 Correlations between the HoNOS and the Brief Psychiatric Rating Scale (BPRS) and Global Assessment Scale (GAS)

Variable	Levels	Time	BPRS	n	GAS	n
Time 1			0.44***	133	0.49***	250
Time 2			0.72***	93	0.71***	149
Change			0.61***	80	0.63***	134
Diagnosis†	Psychosis	T1	0.54***	51	0.52***	79
		T2	0.69***	35	0.70***	48
	Depression	T1	0.55***	39	0.41***	87
		T2	0.67***	25	0.73***	49
Setting†	Acute wards	T1	0.41***	103	0.40***	178
		T2	0.73***	76	0.69***	116
	Day hospital	T1	0.56***	28	0.67***	57
		T2	0.42	15	0.60**	27
Ratings	Team ratings	T1	0.35*	43	0.53***	87
		T2	0.59***	23	0.72***	41
	Single ratings	T1	0.47***	90	0.48***	163
		T2	0.73***	58	0.68***	93

\*p&lt;0.05, \*\*p&lt;0.01, \*\*\*p&lt;0.001 (Spearman rank).

†Number of levels limited by low return of time 2 data.

Table 10 Correlation of social items on HoNOS and comparison scales

HoNOS item	Comparison items	Time 1	n	Time 2	n
Relationships	SPS 5 Intimate relationships	0.23*	124	0.42***	84
	SPS 6 Non-intimate relationships	0.19**	241	0.25**	145
Activities of daily living	SPS 1 Household management	0.16*	239	0.19*	142
	SPS 3 Money management	0.13	233	0.08	136
Social functioning total	SPS total	0.27***	250	0.30***	149

\* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$  (Spearman rank).

ments would enhance the effectiveness of routine quantitative clinical monitoring.

### Discussion

The results indicate that the HoNOS-SMI generally fulfils the requirements of a clinically acceptable outcome scale for routine use in mental health services. HoNOS inter-item correlations were generally low and positive, thus indicating a minimum of redundancy among the 12 items. Principal components analysis extracted four discrete factors from the HoNOS time 1 data, one of which consisted of item 8 on its own. Calculation of item weights showed that the contribution of item 8 to the mean HoNOS total exceeded that of the other items, reflecting the fact that this item is extremely broad conceptually, encompassing a total of nine symptoms. Items 7 and 9 were also scored highly, reflecting the prevalence of depressed mood and interpersonal problems among patients with mental illness. By contrast, item 11 (living conditions) and item 12 (occupation and activities) contributed only 3.3% and 3.4%, respectively, to the mean HoNOS total, indicating that patients were generally scored low on these items. Interviews with HoNOS raters suggested that the information necessary for rating these items, along with the item 10 (activities of daily living) was often unavailable to raters.

HoNOS scores decreased by almost 50% between times 1 and 2, they decreased as the intensity of the setting decreased, time 1 scores for patients detained under the Mental Health Order were 48% higher than those for voluntary patients, and they recorded greater levels of change. Diagnostic groups scored most highly on those items most germane to their illness. Discriminant analysis indicated that the HoNOS had a moderate level of discriminatory power. A closer examination of the discriminant functions showed that some of the failure to discriminate was due to the high scoring of HoNOS item 8 (other mental and behavioural problems), a consequence of the large number of symptoms encompassed by this item. For example, people with eating disorders and disorders of sleep tended to score highly on item 8 and low on all other items, and thus such a profile tended to be allocated to this group. Patients with neurosis/anxiety disorders also scored highly on item 8 but were likely to score moderately on other items such as item 7 (depressed mood). The result is that these cases tended to get classified with a primary diagnosis of depression. Better discriminatory power would be possible if items 5, 11, and 12 were omitted and other items were included to measure independently some of the compo-

nent parts of item 8. Items relating to eating/sleep/stress disorders and anxiety would be useful.

Overall, the HoNOS performed well against other established scales, particularly when sufficient information was available for ratings. When the beginning and end of episodes were examined separately the time 2 data correlated highly but the time 1 correlations did not exceed moderate levels. These differences were partially explained by a degree of variation across diagnostic groups and sites and by information deficits.

Examination of diagnostic groups showed that HoNOS/BPRS associations were strongest at time 1 for the principal diagnostic groups observed (psychosis and depression). Correlations for the other diagnostic groups were low, thus reducing the overall time 1 coefficient to a degree. Examination of settings revealed that the HoNOS/BPRS association at time 1 was weak for one of the acute sites. This site had a particularly large number of raters operating independently without the supervision of a key rater. When data from this site were omitted from calculations the overall time 1 HoNOS/BPRS correlation increased substantially, though not reaching a satisfactory level.

Interviews with HoNOS raters revealed that ratings at time 1 were subject to a deficiency of information. Raters from nine of the 10 sites at which interviews took place were of the view that time 2 ratings were easier to complete and were likely to be more accurate because more information about patients was available at that time. This view was supported by the results from the day hospitals which recorded the strongest correlations at time 1. Day hospital raters tended to delay time 1 ratings for two weeks or more while most acute ward raters scored patients on or soon after admission as part of the initial assessment. As a result, the day hospital raters had more patient information available to them than their acute ward counterparts. In addition, the team rater HoNOS/BPRS correlations at time 1 were higher for these diagnostic groups than for single rater correlations, probably reflecting the greater amount of clinical information generated within a team setting.

Information deficits also accounted largely for the weak performance of the four social items. It was a commonly held view among HoNOS raters that the rating of these items would require the input of someone familiar with the day to day environment of the patients concerned.

In summary, these analyses indicate that the HoNOS is sensitive to change across time and to differences in illness type and severity, and has a sufficient degree of construct validity. HoNOS performs well against established scales when time 2 data are evaluated, but the scale associations at time 1 appear to be adversely affected by a number of factors, particularly information deficits and the quality of on-site supervision of raters. Such factors can be addressed by modifications to operational guidelines. The need for further research on the use of the HoNOS with "other" diagnostic

groups such as anxiety, personality and bipolar disorders is clearly indicated.

#### SUGGESTED MODIFICATIONS TO THE OUTCOME SCALE

While the HoNOS clearly possesses the potential to fulfil the requirements of an outcome scale, its performance may be enhanced by a number of changes.

Modification of item 8 (other mental and behavioural problems) is indicated by its high percentage contribution to scale total scores, its role in reducing the discriminatory power of the scale as a whole, and the dissatisfaction with the item expressed in the course of interviews with HoNOS raters. The type of modification required is a matter for debate but consideration might be given to providing some of the more commonly rated symptoms encompassed by the item, such as anxiety and sleep disturbance, with full item status. A different weighting may be required for such items in order to maintain the balance of the scale.

Statistical analyses revealed a number of problems relating to the social functioning/context items (items 9–12), including low correlations with SPS scores and low scoring weights (items 11 and 12) relative to other HoNOS items. Interviews with HoNOS raters strongly suggested that these problems were the result of a substantial deficit of information. A number of solutions could be considered. Steps could be taken to ensure that items 9–12 are rated by staff who have access to the necessary background information. Alternatively, items 11 and 12 could be excluded from the scale and the rating of items 9 and 10 could be subject to rigorous guidelines that would ensure that they are rated with reference to the necessary information.

Finally, the information derived from correlations and the interviews with raters raised important questions regarding the timing of first ratings. Nine of the 10 sites at which interviews took place stated that the level of confidence in the accuracy of time 1 ratings was considerably less than for those at time 2 because of a relative deficit in information. In the interest of accuracy, ratings of some items could be delayed until the required information is available. However, if the ratings are delayed too long one could argue that they would cease to be useful in outcome measurement.

#### ROUTINE MEASUREMENT IN CLINICAL PRACTICE

Mental health services should ultimately be judged by the evidence of benefit they provide for the people who use the services. This view

has been reinforced by the UK government's vision for assuring quality in the NHS,<sup>11</sup> including the introduction of National Service Frameworks and Clinical Governance. The project experience, in which quantitative clinical ratings were carried out by a large number of raters all in routine practice situations, highlighted a number of factors important for the use of clinical measurement within mental health services. All staff must be introduced to the philosophy of clinical measurement and provided with training and ongoing supervision to ensure that rating practice is standardised and that the quality of rating is maintained. The success of routine clinical monitoring will depend greatly on the level of support provided by management. It is the view of the authors that the practice of outcome measurement should take place within the wider context of quality improvement in which interest in clinical outcomes is part of an organisation wide culture of continuous quality improvement. In this context the HoNOS could form the clinical core of a minimum data set for adult mental illness services. Such routine data capture, supported by a community networked information system, has considerable potential for patient care, the quality agenda, service management, and planning.

The authors gratefully acknowledge the support of the DHSS (NI) Research and Development Office for funding the project, the staff, management and patients of Hospital and Community Trusts who participated in the project, and the Research Unit of the Royal College of Psychiatrists.

Copies of the HoNOS are available from the College Research Unit, 11 Grosvenor Crescent, London SW1X 7EE, UK.

- 1 Delamotho T. *Outcomes in clinical practice*. London: British Medical Journal Publishing Group, 1994.
- 2 Ware IE, Snow KK, Kosinski M, et al. *SF-36 health survey manual and interpretation guide*. Boston: New England Medical Centre, The Health Institute, 1993.
- 3 Wing JK, Curtis RH, Beevor AS. *HoNOS: Health of the Nation Outcome Scales. Report on research and development*. London: College Research Unit, 1996.
- 4 Wing JK, Beevor AS, Curtis RH, et al. Health of the Nation Outcome Scales (HoNOS). *Br J Psychiatry* 1998;172:11–18.
- 5 Royal College of Psychiatrists. *Measuring health outcomes in Mental Health Services. The Health of the Nation Outcome Scale: HoNOS: a guide to the use of HoNOS*. London: College Research Unit, 1995.
- 6 Endicott J, Spitzer RL, Fleiss JL, et al. The Global Assessment Scale. *Arch Gen Psychiatry* 1976;33:766–71.
- 7 Wykes T, Hurry J. Social behaviour and psychiatric disorders. In: Bebbington P, ed. *Social psychiatry: theory, methodology and practice*. New Brunswick, USA: Transaction Publishers, 1991:183–208.
- 8 Overall J, Gorham E. The Brief Psychiatric Rating Scale. *Psychological Reports* 1962;10:799–812.
- 9 Hamilton M. A rating scale for depression. *J Neurol Neurosurg Psychiatry* 1960;3:56–62.
- 10 Hamilton M. (1969) Diagnosis and rating of anxiety. *Br J Psychiatry Special Bulletin* 1969; 76–9.
- 11 NHS. *A first class service, quality in the new NHS*. London: NHS, 1998.