

This is a pre-print of an article submitted to and subsequently accepted for publication in The Journal of Forensic Psychiatry and Psychology © 2007 Taylor and Francis. This is the submitted version of the paper and does not include amendments made following peer review. The Journal of Forensic Psychiatry and Psychology is available online at <http://www.informaworld.com>

HoNOS-secure:

A reliable outcome measure for users of secure and forensic mental health services

Authors:

Geoff Dickens, Research Nurse RMN, BSc(Hons), DipN

Philip Sugarman, Chief Executive & Medical Director, MBChB, MSc, MRC Psych

Lorraine Walker, Head of Healthcare Governance Dept, RMN, MA, PGCE

Department/ Address for correspondence:

Research Department

St Andrew's Hospital

Northampton

NN1 5DG

ABSTRACT

The Health of the Nation Outcome Scales for users of Secure and Forensic Services (HoNOS-*secure*) comprises amended versions of the original 12 HoNOS scales, and seven security scales. HoNOS-*secure* tracks clinical outcome, including ongoing security needs. We investigated the interrater reliability of HoNOS-*secure*. Sixty inpatients were rated independently by two clinicians. There were 34 raters in total. Intraclass correlation coefficients (ICC's) for 11 of the 12 HoNOS items indicated fair to substantial consistency between raters. ICC's for six of the seven security scales indicated moderate, substantial, or better, agreement; one item indicated fair consistency. Most security scale and HoNOS items were measured consistently by different raters, indicating HoNOS-*secure* is a reliable outcome measure when used in routine clinical practice.

Keywords:

Outcome measures, Secure services, Psychometry, Mentally Disordered Offenders.

INTRODUCTION

Planned expansions will have increased by 50% the number of medium-secure mental health inpatient placements in England and Wales to over 3000 since 2001 (Jaycock & Bamber, 2001). In addition, specialist forensic community mental health teams (Mohan *et al*, 2004) have recently emerged, in addition to the existing range of high and low security inpatient provision, and prison-based mental health services.

Service-users commonly have a mental illness and a history of criminal behaviour (e.g. Coid, 2001) though this broad definition masks the underlying range of problems and patient characteristics: mental illness including schizophrenia, personality disorder, offending behaviour including violent and sexual crimes, and behavioural disturbance including self-harm or co-morbid substance misuse. Despite this increase in the scale and breadth of secure and forensic services, there is currently no nationally used outcome measure for users of secure services.

The Health of the Nation Outcome Scales (HoNOS) were initially developed for use with individuals with mental health problems in the general population (Wing *et al*, 1998; Bebbington *et al*, 1999). HoNOS-*secure* was originally conceived as HoNOS-MDO (Mentally Disordered Offenders) in collaboration with the Royal College of Psychiatrists College Research Unit. The current version (HoNOS-*secure* version 2.0) was devised by two of the current authors (PS, LW) and has undergone qualitative development and substantial revision, using consultation and case vignettes in order to establish face validity.

Interrater reliability measures the degree to which two independent raters agree on score items for individuals. It is “especially important [to establish interrater reliability] when measuring devices are new” (Portney & Watkins, 1993). The aim of this study, therefore, was to establish whether HoNOS-*secure* items can claim to have interrater reliability in routine clinical practice.

METHOD

Setting

The study took place in three mental health services (secure adult mental health, secure adult learning disability, and secure geriatric mental health) at St Andrew's Hospital, Northampton, a specialist independent sector provider of inpatient mental health services.

Raters

Raters were 34 healthcare professionals employed by the three services, including psychiatrists, psychiatric nurses, psychologists and occupational therapists. All raters knew the patients concerned well, either in their position as Responsible Medical Officer, Care Programme Approach coordinator or as a senior qualified nurse. All raters had received training in use of the HoNOS-*secure* scales, and all were familiar with their routine clinical use.

Participants

Sixty inpatients with mental disorder aged 18 to 70 years. All were current inpatients in low or medium secure wards in one of the three services included in the study.

Design and materials

Each patient was rated independently by two raters. No single rater assessed more than six participants. Assessments were made separately by the two raters in the course of routine clinical practice. Raters had access to patient's medical notes

including any current risk assessments but did not interview them for the study. All ratings were coordinated by author GD independently from the other two authors.

HoNOS-*secure* consists of the 12 original HoNOS items (see *Table 1*), modified to account for secure settings, plus seven security scale items measuring need for secure measures, including dangerousness and need for physical, relational and procedural security measures (see *Table 2*). Whilst HoNOS items are rated retrospectively for observed problem behaviour, the security scales are rated prospectively for the period “in the near future, including if living unsupported in the community”. HoNOS-*secure* is accompanied by a glossary to aid completion. All items, including secure scale items, are rated on a five point scale (0 to 4), each scalar point having a brief narrative description in the glossary to provide an anchor point for the score. Ratings were entered into SPSS 11.0 for analysis.

Statistical Analysis

The intraclass correlation coefficient (ICC) provides an assessment of interrater reliability by comparing the amount of variation between individual raters with overall variance. This is the appropriate test to use when multiple raters are involved and is a measure of reliability equivalent to weighted kappa (Fleiss & Cohen, 1973). Landis & Koch (1977) suggest that a test of the reliability of an instrument is provided by the following rubric: 0.21 – 0.40 = fair, 0.41 – 0.60 = moderate, 0.61 – 0.80 = substantial, 0.81 – 1.00 = almost perfect. Additionally, calculation of the test statistic (P), and 95% Confidence Intervals (CI) for the ICC, indicates whether pairs of raters agreed at a level which is unlikely to be explicable by chance alone.

RESULTS

Security scale items

Table 1 displays mean scores for each security scale item (range 1.30 – 2.18). All means lie within a range which reflects the nature of services in the current study setting i.e. low and medium secure facilities. Highest standard deviations were for items A (Harm to others) and B (Harm to self), suggesting there are quite wide variations among patients in this study on these behaviours. Standard deviations for other items are relatively low suggesting similar level of need among participants.

Table 1
about here

Calculation of test statistics for ICC's (see *Table 2*) indicated that, for all items, rating was consistent between raters at levels significantly greater than chance. Interrater reliability was substantial or better for items A to E, and moderate for item G. The ICC for Item F (Harm from others) suggests only fair reliability between raters.

Table 2
about here

HoNOS items

Table 3 displays mean scores for each HoNOS scale item. Mean scores are generally low (range 0.07 – 1.47), possibly suggesting that the psychiatric symptomatology of patients in the study is well managed. On the items that most clearly duplicate each other (secure scale item B 'Self-harm' and HoNOS Scale 2 'Non-accidental self-injury') ratings differed significantly for the same patients, with scores on HoNOS being significantly lower (mean = 0.33, s.d. = 0.63) than scores on the secure scale (mean = 1.46, s.d. = 1.06, d.f. = 119, $P < .0001$), suggesting that the instruction to rate retrospectively observed behaviour on HoNOS scales and "near future" behaviour on

Table 3
about here

the secure scales is understood. As with previous HoNOS- related studies (e.g. Fleminger *et al*, 2005), scores for alcohol/ drug use are extremely low, perhaps reflecting the secure study setting and the consequent perceived lack of access to alcohol or other illicit substances. *Table 4* displays ICC values for HoNOS items. Interrater reliability was moderate to substantial for nine of the twelve items, and fair for the remaining three items. Curiously, item 7 (Depression), did not appear to be a consistently rated item as indicated by the presence of zero in the 95% CI.

Table 4 about here

DISCUSSION

We aimed to evaluate the interrater reliability of HoNOS-*secure* by asking pairs of raters to assess participants in the course of routine clinical practice. Intraclass correlation coefficients (ICC's) suggested that raters agreed at levels significantly above chance for all seven security scales and all of the 12 HoNOS scales. An ICC of 0.39 and a relatively wide 95% Confidence Interval (0.43) suggested only fair agreement on the security scale item 'Harm from others', suggesting that this item is difficult to accurately assess quantitatively. Information about vulnerability or potential for exploitation may be more usefully viewed qualitatively. Security scale item G (Need for risk management procedures) was rated moderately consistently but the wide confidence interval (0.37) suggests that review of each scalar narrative point may clarify understanding. Items A to E all showed substantial or better interrater reliability with relatively narrow 95% Confidence Intervals, and on this evidence can be claimed to be reliable measures between raters in routine clinical practice.

ICC's for all HoNOS items indicated that raters agreed at levels significantly above chance. As with previous studies (e.g. Fleminger et al, 2005) scores for alcohol/ drug misuse were extremely low, indicating that raters score on the basis of current access to substances in a secure setting rather than potential for misuse in the community. Items 11 (Problems with living conditions) and 12 (Problems with occupations and activities) can be described as having fair consistency. These items were originally developed for community patients and have consistently been the items with lowest interrater reliability in other HoNOS-family studies (e.g. Trauer *et al*, 1999), and are identified as problematical in multiple studies in a recent comprehensive review of

reliability and validity of HoNOS family members (Pirkis *et al*, 2005). Rating of depressed mood also appeared to lack consistency in this study, and this is also not new (e.g. Bebbington *et al*. 1999) and may reflect the low numbers of psychiatrists with specialist skills in detecting depressive symptoms acting as raters in this study.

Potential criticisms

The study was undertaken in a single centre, a specialist independent sector hospital and generalisation to other settings cannot be assumed. However, independent sector providers represent at least a third of medium-secure placements (Jaycock & Bamber, 2001), and clinical staff generally have experience of working in various settings including the NHS. Although this was a single site study it should be emphasised that participants were drawn from 3 separately-registered Independent Hospitals on the same site (secure adult mental health service, secure adult learning disability services, and secure geriatric mental health services), and that users are from many parts of the country. HoNOS-*secure* scales are intended to work alongside other HoNOS-family members. We used a range of professionals as raters, and there is some evidence to suggest that different professions view issues of risk and outcome differently. On the other hand, our study reflected real life practice in that outcome scales are routinely completed by professionals from varying disciplines. All of our raters had received training in the use of HoNOS-*secure* scales, which may have improved reliability, though it has previously been noted that training staff in HoNOS rating may have no value (Brooks, 2000).

Implications of the study

This study of interrater reliability suggests that HoNOS-*secure* is a promising outcome measure for users of forensic and secure services. There may be a need to re-examine the secure scale item measuring potential for harm from others. We are not currently aware of any assessment tool that claims to measure this construct with both reliability and validity, and its use should be treated cautiously. There is clearly a need to establish HoNOS-*secure*'s validity as a priority through further research, and also to undertake further tests of reliability, including test-retest reliability. Further studies will also need to include users of community forensic mental health services and high secure hospitals. For most items, HoNOS-*secure* proved to be a reliable measure of outcome for users of the three forensic/ secure mental health services in this study during the course of routine clinical practice.

REFERENCES

- Bebbington, P., Brugha, T., Hill, T. *et al.* (1999) Validation of the Health of the Nation Outcome Scales. *British Journal of Psychiatry*, **174**, 389–394.
- Brooks, R. (2000) The reliability and validity of the Health of the Nation Outcome Scales: validation in relation to patient derived measures. *Australian and New Zealand Journal of Psychiatry*, **34**, 504-511.
- Coid, J., Kahtan, N., Gault, S., Cook, A. & Jarman, B. (2001) Medium secure forensic psychiatry services: Comparison of seven English health regions. *British Journal of Psychiatry*, **178**, 55 – 61.
- Fleiss, J.L. & Cohen, J. (1973) The equivalence of weighted kappa and the intraclass correlation coefficient as measures of reliability. *Educational and Psychological Measurement*, **33**, 613–619.
- Fleminger, S., Leigh, E. Eames, P. *et al* (2005) HoNOS-ABI: a reliable outcome measure of neuropsychiatric sequelae to brain injury? *Psychiatric Bulletin*, **29**, 53–55.
- Jaycock & Bamber (2001) Mental Health Services: On the lookout. *Health Service Journal*, **111 (5780)**, 26-27.
- Landis, J.R. & Koch, G.G. (1977) The measurement of observer agreement for categorical data. *Biometrics*, **33**, 159–174.

Mohan, R., Slade, M., & Fahy, T.A. (2004) Clinical characteristics of community mental health services. *Psychiatric Services*, **55(11)**, 1294–8.

Pirkis, J.E., Burgess, P.M., Kirk, P.K. *et al* (2005) A review of the psychometric properties of the Health of the Nation Outcome Scales (HoNOS) family of measures. *Health and Quality of Life Outcomes*, **3(76)**. Full text available open access online at Biomed Central <http://www.hqlo.com/content/3/1/76> (date accessed 27th March, 2006).

Portney, L.G. & Watkins, M.P. (1993) *Foundations of clinical research: Applications to practice*. Stamford, Ct: Appleton & Lange.

Trauer, T., Callaly, P., Hantz, P. *et al* (1999) Health of the Nation Outcome Scales: Results of the Victorian field trial. *British Journal of Psychiatry*, **174**, 380–388.

Wing, J.K., Beevor, A.S., Curtis, R.H. *et al* (1998) Health of the Nation Outcome Scales (HoNOS). Research and development. *British Journal of Psychiatry*, **172**, 11-18.

Table 1: Mean (s.d.) scores on each 'secure scale' item (0 - 4) and total score N = 60

Item	Descriptor	Mean (s.d.)
A	Harm to adults or children	1.97 (1.13)
B	Self-harm (deliberate or accidental)	1.37 (1.13)
C	Need for building security to prevent escape	2.18 (0.65)
D	Need for a safely staffed living environment	1.93 (0.58)
E	Need for escort on leave	1.40 (0.74)
F	Harm from others	1.30 (0.94)
G	Need for risk management procedures	1.68 (0.62)
TOTAL SCORE		11.78 (3.98)

Table 2: Intraclass correlation values for all secure scale items N = 60

Item	Descriptor	ICC (95% CI)	P
A	Harm to adults or children	0.64 (0.46 – 0.77)	P<.0001
B	Self-harm (deliberate or accidental)	0.67 (0.53 – 0.80)	P<.0001
C	Need for building security to prevent escape	0.88 (0.81 – 0.93)	P<.0001
D	Need for a safely staffed living environment	0.70 (0.55 – 0.81)	P<.0001
E	Need for escort on leave	0.88 (0.81 – 0.93)	P<.0001
F	Harm from others	0.39 (0.15 – 0.58)	P<.001
G	Need for risk management procedures	0.53 (0.32 – 0.69)	P<.0001

Table 3: Mean (s.d.) scores on each HoNOS item (0 – 4) N = 60

Item	Descriptor	Mean (s.d)
1	Antisocial	1.35 (1.14)
2	Self-harm	0.33 (0.63)
3	Alcohol/ Drugs	0.10 (0.40)
4	Cognitive	0.98 (1.01)
5	Physical	1.27 (1.23)
6	Psychotic	0.85 (1.07)
7	Depression	1.10 (0.91)
8	Other	1.33 (1.16)
9	Relationships	1.67 (0.98)
10	Activities of Daily Living	1.68 (1.33)
11	Living conditions	0.65 (0.68)
12	Occupations and activities	0.75 (0.87)

Table 4: Intraclass correlation values for all HoNOS items N = 60

Item	Descriptor	ICC (95% CI)	P
1	Antisocial	0.89 (0.82 – 0.93)	P<.0001
2	Self-harm	0.66 (0.49 – 0.78)	P<.0001
3	Alcohol/ Drugs	0.96 (0.93 – 0.98)	P<.0001
4	Cognitive	0.76 (0.63 – 0.85)	P<.0001
5	Physical	0.65 (0.48 – 0.77)	P<.0001
6	Psychotic	0.81 (0.70 – 0.88)	P<.0001
7	Depression	0.24 (-0.01 – 0.46)	P<.05
8	Other	0.65 (0.48 – 0.78)	P<.0001
9	Relationships	0.52 (0.32 – 0.68)	P<.0001
10	Activities of Daily Living	0.88 (0.80 – 0.93)	P<.0001
11	Living conditions	0.39 (0.16 – 0.59)	P<.001
12	Occupations and activities	0.37 (0.13 – 0.57)	P<.01