BMJ Open

Consistency of the Health of the Nation Outcome Scales (HoNOS) at inpatient-to-community transition

Wei Luo,[1] Richard Harvey,[2] Truyen Tran,[3] Dinh Phung,[3] Svetha Venkatesh,[3] Jason P Connor[4]

For numbered affiliations see end of article.

Correspondence to
Dr Wei Luo;
wei.luo@deakin.edu.au

ABSTRACT

Objectives: The Health of the Nation Outcome Scales (HoNOS) are mandated outcome-measures in many mental-health jurisdictions. When HoNOS are used in different care settings, it is important to assess if setting specific bias exists. This article examines the consistency of HoNOS in a sample of psychiatric patients transitioned from acute inpatient care and community centres.

Setting: A regional mental health service with both acute and community facilities.

Participants: 111 psychiatric patients were transferred from inpatient care to community care from 2012 to 2014. Their HoNOS scores were extracted from a clinical database; Each inpatient-discharge assessment was followed by a community-intake assessment, with the median period between assessments being 4 days (range 0–14). Assessor experience and professional background were recorded.

Primary and secondary outcome measures: The difference of HoNOS at inpatient-discharge and community-intake were assessed with Pearson correlation, Cohen's κ and effect size.

Results: Inpatient-discharge HoNOS was on average lower than community-intake HoNOS. The average HoNOS was 8.05 at discharge (median 7, range 1–22), and 12.16 at intake (median 12, range 1–25), an average increase of 4.11 (SD 6.97). Pearson correlation between two total scores was 0.073 (95% CI −0.095 to 0.238) and Cohen's κ was 0.02 (95% CI −0.02 to 0.06). Differences did not appear to depend on assessor experience or professional background.

Conclusions: Systematic change in the HoNOS occurs at inpatient-to-community transition. Some caution should be exercised in making direct comparisons between inpatient HoNOS and community HoNOS scores.

Strengths and limitations of this study

- Consistency of Health of the Nation Outcome Scales (HoNOS) scores were based on data linkage between different facilities with largely overlapping assessment timeframes.
- The data were routinely collected in a naturalistic setting with diversity in consumer needs and clinician background, reflecting typical use of the HoNOS.
- The study is limited by its retrospective nature.
- Patients in this study were primarily treated for psychosis/schizophrenia and findings many not be generalisable to other mental health conditions.

Zealand jurisdictions, also gaining popularity in a number of other countries.[1–6] The HoNOS serve not only as a tool to evaluate patient functioning, but also as an index for comparative service performance and cost-effectiveness analyses.[7] Given mandated application of the HoNOS, it is a timely question to ask whether HoNOS can be interpreted consistently across different care settings. In this study, we examined stability of the HoNOS scores in a sample of public health service psychiatric consumers transitioning from inpatient to community care.

HoNOS background: Reliable measures of patient mental response to treatment is central to care planning and quality assurance. Symptom rating scales such as the Beck Depression Inventory (BDI[8]), State-Trait Anxiety Inventory (STAI[9]) and Positive and Negative Syndrome Scale (PANSS[10]) are well-recognised individual assessment tools that reliably measure patient mental health symptoms. Many symptom rating scales do not comprehensively assess overall patient health and social functioning. These domains are typically seen by health and government services as more critical in evaluating patient outcomes.[11] To address the need for a more standardised and systematic approach at

INTRODUCTION

The Health of the Nation Outcome Scales (HoNOS) are now routinely applied as a mandatory mental health outcome measure in the UK and many Australian and New

measuring health outcomes for mental health patient functioning, the HoNOS was developed.[12] [13] The HoNOS has 12 items covering behaviour, impairment, symptoms and social functioning domains.[13]

Early HoNOS psychometric validation studies were unconvincing. A 6-week pre–post administration analysis of the HoNOS by Bebbington et al[14] concluded that there were "serious problems in using the instrument as a routine measure of clinical status in busy psychiatric services." (p. 389). They found that performance of the instrument was likely to be related to the assessors' level of training and skill, which reflected clinical reality in most busy mental health services. Similar evaluations around this time also failed to provide compelling support for the widespread adoption of the HoNOS, primarily on psychometric grounds.[15] [16] One of the key concerns was poor reliability.[14] [16] Despite no modification to the scale itself, subsequent reliability testing has provided additional support for its utility in clinical practice, based on larger and more heterogeneous samples, and arguably more sophisticated statistical approaches to validation. The reluctance of some clinical staff to fully adopt routine HoNOS assessments has been reported[17] and there remain some concerns about the utility of the scale to inform clinical, as opposed to population level change.[18]

In a comprehensive systematic narrative review of HoNOS validation studies, Pirkis et al[1] concluded that the scale had 'good' validity (ie, it accurately measures what it was designed to measure) and had 'adequate' test–retest reliability (ie, consistently measures functioning at different time points) and inter-rater reliability (high correlations between different assessors scores for the same patient). Pirkis et al[1] did note that few HoNOS studies have formally conducted test–retest reliability (n=3), and one-third of the 12 items consistently performed poorly on this psychometric index. For example, in a sample of 100 inpatients, Orrel et al's[19] psychometric evaluation found the 1 week test–rest reliability of HoNOS items ranged from 0.33 to 0.80 (mean 0.57), with only two items exceeding the accepted 0.7 cut-off.[20] Pirkis et al's[1] review also found inter-rater agreement has been identified as poor on at least 50% of the items. Orrel et al's reported inter-rater item reliability ranged from −0.03 to 0.65 (mean 0.395), with only one of the items exceeding the recognised cut-off for 'good' (0.60–0.74) reliability.[21] In a sample of 100 older psychiatric inpatients (average age=77 years), all but two items exceed minimum test–retest or inter-rater reliability.[22] Brooks[3] also found 'limited' inter-rater reliability (0.50–0.65) and concluded that "the HoNOS not be implemented as a major outcome tool, until the reliability and validity of the HoNOS is clearly established." (pp. 509–510).

HoNOS at care transition: In the Australian state of Victoria, when a consumer is discharged from an inpatient unit and subsequently admitted to a community residential unit, both an inpatient 'discharge' HoNOS and a subsequent community 'intake' HoNOS are completed. These data provide two assessments within a practically overlapping timeframe. Examining these data can provide important evidence on the consistency of HoNOS across mental healthcare settings.

METHODS

Subjects

The study population includes all mental health patients from a regional mental health service who were transferred from an acute psychiatric unit to a number of community teams and received both inpatient discharge HoNOS and subsequent community intake HoNOS between July 2012 and June 2014. There were 70 males (mean age 38, SD 12.2) and 41 females (mean age 41.5, SD 13.3). The majority of the sample were being treated for schizophrenia, schizotypal or delusional disorders (n=63).

Procedure

HoNOS were administered to inpatients on discharge by clinically trained inpatient staff, and within a maximum of 14 days by clinically trained community-based mental health staff. The median time between inpatient and community HoNOS assessments was 4 days, with 95% administered within 12 days.

Measures

The HoNOS[12] [13] consists of 12 items, each ranging from 0 (no problem) to 4 (severe problem). As recommended by the scale developers, the 12 items are scored into four subscales:[12] Behaviour (items 1–3), Impairment (items 4–5), Symptoms (items 6–8), and Social (items 9–12). The sum of the 12 items provides a total score (range of 0–48). HoNOS can be interpreted at the item, subscale or total score level. The rating period is generally the preceding 2 weeks for inpatients at admission, for hospital outpatients and for all clients of community-based services.[23] [24] The exception is at discharge from acute inpatient care, in which case the rating period should generally be the preceding 3 days and not collected if length of stay is less than 72 h. A community intake rating is collected at the allocation of case manager or first care plan.[25] Therefore, the second assessment period is expected to overlap with the initial assessment period.

Statistical analysis

Both inpatient and community HoNOS scores for each patients were paired and the average difference computed. Paired exact Wilcoxon signed-rank test was performed at the three scoring levels of the HoNOS (item, substance, total score). R package exactRankTests was used for computing this test statistics (exactRankTests: Exact Distributions for Rank and Permutation Tests

[program], Torsten Hothorn, Kurt Hornik; 2013). Both Pearson correlation and Cohen's κ were calculated. χ^2 Analyses examined if professional background and level of experience contributed to differences in HoNOS scores between inpatient and community assessments.

RESULTS

Total HoNOS score

The average inpatient discharge HoNOS score was 8.05 (median 7, range 1–22) and the average community intake HoNOS was 12.16 (median 12, range 1–25). Cronbach's α for discharge score was 0.7 (95% CI 0.6 to 0.79); α for intake score was 0.69 (95% CI 0.6 to 0.79). Pearson correlation between the two scores was 0.073 (95% CI −0.095 to 0.238), and Cohen's κ was 0.02 (95% CI −0.02 to 0.06), showing poor consistency among two scores. (The numbers are slightly higher when two assessments were performed within 4 days, where Pearson correlation was 0.14 (95% CI −0.11 to 0.37) and Cohen's κ was 0.07 (95% CI −0.01 to 0.14)). On average, the total HoNOS score increased by 4.11 (SD 6.97) from an inpatient discharge to a community intake. The change of the total score reached the threshold of 4–8 points for clinically relevant change proposed by Parabiaghi et al[4 26] and validated by Egger et al.[27] Exact Wilcoxon signed-rank test for matched data returned p value<0.001, indicating that the HoNOS increase from inpatient discharge and community intake is statistically significant. Cohen's d was 0.82 (95% CI 0.58 to 1.07), indicating a large effect size (of difference). The power based on paired t test was 1.

Schizophrenia, schizotypal and delusional disorders (F2x) were the most common diagnoses (56.8%), followed by mood disorders (F3x, 10.8%). Given the low frequency of other diagnoses (5 cases in adjustment disorders, acute stress reaction, anxiety disorder, substance intoxication and dementia), an 'other' category was used and it also included all episodes with no diagnosis recorded in the database. When diagnostic groups were considered, F2x and 'Other' diagnoses demonstrated significant differences between inpatient and community assessments (4.42 and 4.16, respectively). The differences between assessments also reached the Paragbiaghi's threshold for clinically relevant change. There were no significant differences reported in patients with mood disorders (table 1).

HoNOS subscales and individual items

The increase in scores between assessments was also evidenced in HoNOS subscales (table 2) and individual score items (table 3). Among Wing's four-factor subscales,[12] only the change in Social did not have a p value<0.001 and a t test power of nearly 1. Using Speak-Hay-Muncer's four-factor model[28] and Lovaglio's six-item subscale[29] also resulted in statistically significant change between assessments. The only score items that did not reach statistical significance were item 11: Problems with living conditions and item 12: Problems with occupation and activities.

Clinician profession and experience

To examine if clinical professional background and level of experience contributed to differences in HoNOS scores, the proportion of staff profession/experience between Inpatient clinicians and community clinicians was examined. Of the patients in the sample, 98 (82%) were assessed by experienced inpatient nurses and 94 (79%) were assessed by experienced community nurses, indicating equivalence across services (experienced nurses vs others by community vs hospital χ^2=0.2425, p=0.622). The remaining staff included psychologists (n=15), less experienced nurses (n=2) and students (n=11). Two analysis of variance were conducted which examined potential differences in HoNOS scores within each service, by profession and experience. (The score changes at transition were mostly normally distributed, with Shapiro-Wilk test returning a p value of 0.20). No evidence supported potential bias in score changes introduced by experience levels at discharge (F(2, 109) =0.11, p=0.90) or at intake (F(1107)=0.70, p=0.50).

DISCUSSION

The HoNOS has widespread, mandated use in many public health mental health systems including the UK and Australia. It is designed to measure health outcomes for mental health patient functioning at both a clinical and population level. It has been used as key index for comparative service performance and cost-effectiveness analyses,[7] and therefore can be applied as a lever for resource allocation. Given the scope of HoNOS roll out, reliability of measurement is critical. Perhaps surprisingly, the reliability of the HoNOS has received little

Table 1 Change of the total HoNOS scores by diagnosis

Diagnosis	Average (range)		Change	p Value	Pearson correlation (95% CI)	Cohen's κ (95% CI)
	Discharge	**Intake**				
F2x: schizophrenia, schizotypal and delusional disorders (n=63)	7.58 (1–20)	12.0 (3–25)	4.42	<0.001	0.07 (−0.18 to 0.32)	−0.01 (−0.06 to 0.03)
F3x: mood (affective) disorders (n=12)	7.75 (1–18)	10.04 (1–19)	2.29	0.189	−0.15 (−0.69 to 0.49)	0.06 (−0.10 to 0.22)
Other diagnoses (n=36)	8.81 (1–22)	12.97 (2–24)	4.16	<0.001	0.11 (−0.14 to 0.34)	0.04 (−0.02 to 0.11)

HoNOS, Health of the Nation Outcome Scales.

Table 2 Changes of the four HoNOS subscales defined in [12] and HoNOS-6 defined in ref. [29]

HoNOS subscales	Average (range)		Change	Cohen's d (95% CI)	p Value (power)	Pearson correlation (95% CI)	Cohen's κ (95% CI)
	discharge	Intake					
Wing et al's original four-factor subscales							
Behaviour	0.84 (0–5)	2.0 (0–7)	1.16	0.75 (0.50 to 1.0)	<0.001 (1)	0.24 (0.07 to 0.39)	0.15 (0.07 to 024)
Impairment	1.33 (0–5)	2.10 (0–6)	0.77	0.50 (0.26 to 0.74)	<0.001 (1)	0.43 (0.28 to 0.56)	0.08 (−0.01 to 0.17)
Symptoms	2.52 (0–7)	3.86 (0–10)	1.34	0.75 (0.50 to 0.99)	<0.001 (1)	0.15 (−0.02 to 0.31)	0.07 (0 to 0.13)
Social	3.35 (0–12)	4.20 (0–13)	0.84	0.33 (0.09 to 0.57)	0.006 (0.97)	0.18 (0 to 0.33)	0.02 (−0.04 to 0.08)
Speak et al's four-factor model							
Personal well-being	2.50 (0–10)	3.69 (0–13)	1.19	0.50 (0.26 to 0.74)	<0.001 (1)	0.31 (0.15 to 0.45)	0.05 (−0.02 to 0.12)
Emotional well-being	1.59 (0–6)	2.95 (0–11)	1.36	0.80 (0.55 to 1.05)	<0.001 (1)	0.27 (0.11 to 0.42)	0.06 (−0.01 to 0.14)
Social well-being	2.23 (0–9)	3.47 (0–12)	1.24	0.56 (0.31 to 0.80)	<0.001 (1)	0.09 (−0.08 to 0.25)	0.08 (0 to 0.15)
Severe disturbance	1.15 (0–4)	1.88 (0–7)	0.74	0.56 (0.31 to 0.80)	<0.001 (1)	0.26 (0.10 to 0.42)	0.13 (0.03 to 0.22)
HoNOS-6 subscales							
Lovaglio and Monzani's	4.21 (0–13)	6.08 (0–16)	1.87	0.58 (0.34 to 0.83)	<0.001 (1)	0.19 (0.02 to 0.35)	0.02 (−0.03 to 0.08)

HoNOS-6, Health of the Nation Outcome Scales-six item.

empirical attention. Within a publically funded mental health system, this study identified poor total and sub-score consistency across treatment settings. The median delay between an inpatient-discharge HoNOS and the following community-intake HoNOS was 4 days, with 95% occurring within 12 days (range 0—14 days). All individual items on the HoNOS on this sample performed poorly. No systematic differences in the assessors' professional background or level of clinical experience were identified.

The mean difference between scores following inpatient discharge to community assessment was 4 points, or approximately a one-third decrease in clinician assessed patient functioning. This reached the 4 point cut-off of clinically relevant change suggested by Parabiaghi et al[4 26] and later corroborated by Egger et al.[27] Using the thresholds proposed by Parabiaghi et al[26] (13 between moderate and severe illness; 10 between moderate and mild illness), among 81 consumers deemed mild at hospital discharge, 14 would have had changed to moderate at community intake and 32 would have had changed to severe.

Differences in HoNOS scores may also reflect the severity of the population under investigation. Orrell et al (1999) did show average HoNOS inpatient scores to be higher than patients assessed in outpatient or community setting. Inpatient status is typically indicated for patients that cannot be managed as an outpatient or community patients, due to the severity or complexity of the condition. It is important to determine if these observed differences reflect true variation across inpatient and community functioning, or vulnerability in reliability of measurement. Counter-intuitively, in the current study, patients assessed in the community were rated as functioning significantly more poorly than when assessed as an inpatient with overlapping assessment windows (median 4-day interval).

The recommended timeframe for the HoNOS assessment may provide some additional clues to poor reliability. At discharge, the period of assessment is recommended to cover the past 72 h.[24] In the community, the recommendation assessment period is within the past 2 weeks.[24] The most likely clinical reason for discharge is improved functioning and/or increased symptom stability. Patients who have improved functioning over the past 3 days are among those likely to be considered for inpatient discharge. When assessed in the community, a 2-week period is reviewed which may encapsulate broader functioning difficulties. Functioning may have also deteriorated due to less intensive monitoring and patient management. However, given the two review periods overlap and higher community ratings are consistent across all domains, differences point towards measurement problems.

The findings should be interpreted in light of some limitations. Although previous reliability studies have typically occurred in similar or smaller sample sizes,[3 19 22] a larger and more heterogeneous psychiatric sample

Table 3 Changes of the 12 HoNOS score items

HoNOS score items	Average (range)			p Value	Pearson correlation (95% CI)	Cohen's κ (95% CI)
	Discharge	Intake	Change			
1. Overactive, aggressive, disruptive behaviour	0.31 (0–2)	0.73 (0–4)	0.42	<0.001	0.34 (0.19 to 0.49)	0.15 (0.03 to 0.27)
2. Non-accidental self-injury	0.13 (0–2)	0.44 (0–4)	0.31	<0.001	0.24 (0.08 to 0.39)	0.13 (0 to 0.27)
3. Problem-drinking or drug-taking	0.41 (0–3)	0.83 (0–4)	0.43	<0.001	0.31 (0.14 to 0.45)	0.19 (0.08 to 0.31)
4. Cognitive problems	0.71 (0–3)	1.03 (0–4)	0.32	<0.001	0.24 (0.08 to 0.40)	0.05 (−0.06 to 0.17)
5. Physical illness or disability problems	0.61 (0–3)	1.07 (0–4)	0.45	<0.001	0.59 (0.46 to 0.69)	0.26 (0.15 to 0.37)
6. Problems associated with hallucinations and delusions	0.88 (0–3)	1.21 (0–4)	0.33	<0.001	0.31 (0.15 to 0.45)	0.24 (0.13 to 0.36)
7. Problems with depressed mood	0.76 (0–3)	1.03 (0–4)	0.28	<0.001	0.29 (0.13 to 0.44)	0.09 (−0.04 to 0.21)
8. Other mental and behavioural problems	0.86 (0–3)	1.61 (0–4)	0.68	<0.001	0.26 (0.07 to 0.43)	0.11 (0 to 0.22)
9. Problems with relationships	1.11 (0–4)	1.70 (0–4)	0.59	<0.001	0 (−0.18 to 0.16)	0.02 (−0.07 to 0.12)
10. Problems with activities of daily living	0.88 (0–3)	1.34 (0–4)	0.46	<0.001	0.34 (0.19 to 0.48)	0.08 (−0.03 to 0.19)
11. Problems with living conditions	0.55 (0–3)	0.63 (0–4)	0.08	0.27	0.19 (0 to 0.36)	0.02 (−0.12 to 0.15)
12. Problems with occupation and activities	0.55 (0–3)	0.52 (0–3)	−0.05	0.70	−0.04 (−0.23 to 0.14)	−0.05 (−0.2 to 0.1)

12 HoNOS, Health of the Nation Outcome Scales-12 item.

across a broader age range would provide a higher powered study to more robustly detect associations between testing periods and assessors. Patients in this study were primarily treated for psychosis/schizophrenia and findings many not be generalisable to other mental health conditions. We cannot confirm all assessors received the same level of training in the HoNOS. Given training differences have been previously detected,[14] there may have been systematic biases associated with either inpatient or community assessors. The interval period for test–rest reliability ranged from 0 to 14 days, with a median of 4 days. The timeframe is consistent with previous HoNOS reliability studies that have retested up to 6 weeks.[14] The modest variation in retesting timeframes in the current study is not psychometrically desirable, however the clear pattern of results demonstrating an upward assessment of severity across domains provides some confidence that timeframe variations did not confound findings. Finally, the score consistency was also subject to potential time biases. A central purpose of this study was to assess the ecological reliability of the HoNOS in a busy clinical mental health service. A more robust design would video record patient assessments and responses (at one point in time) and has multiple clinicians blind to the original assessment rescore the scale.

In summary, limited support was found to support consistent interpretation of the HoNOS at different settings in this psychiatric patient sample. Recognising the scale has already been widely mandated in clinical practice, further larger scale testing is considered critical to ensure these findings are not simply a result of sample or assessor bias.

Author affiliations
[1]School of Information Technology, Deakin University, Geelong, Victoria, Australia
[2]Mental Health Services, Barwon Health, School of Medicine, Deakin University, Geelong, Victoria, Australia
[3]Centre for Pattern Recognition and Data Analytics, Deakin University, Australia, Geelong, Victoria, Australia
[4]Discipline of Psychiatry and Centre for Youth Substance Abuse Research, The University of Queensland, Brisbane, Queensland, Australia

REFERENCES
1. Pirkis JE, Burgess PM, Kirk PK, *et al.* A review of the psychometric properties of the Health of the Nation Outcome Scales (HoNOS) family of measures. *Health Qual Life Outcomes* 2005;3:76.
2. Glover G, Knight S, Melzer D, *et al.* The development of a new minimum data set for specialist mental health care. *Health Trends* 1997;29:48–51.
3. Brooks R. The reliability and validity of the health of the nation outcome scales: validation in relation to patient derived measures. *Aust N Z J Psychiatry* 2000;34:504–11; discussion 20-1.

4. Parabiaghi A, Rapisarda F, D'Avanzo B, *et al*. Measuring clinical change in routine mental health care: differences between first time and longer term service users. *Aust N Z J Psychiatry* 2011;45:558–68.

5. Kortrijk HE, Mulder CL, Drukker M, *et al*. Duration of assertive community treatment and the interpretation of routine outcome data. *Aust N Z J Psychiatry* 2012;46:240–8.

6. Theodoridou A, Jaeger M, Ketteler D, *et al*. The concurrent validity and sensitivity of change of the German version of the health of the nation outcome scales in a psychiatric inpatient setting. *Psychopathology* 2011;44:391–7.

7. Slade M, Byford S, Barrett B, *et al*. Alternatives to standard acute in-patient care in England: short-term clinical outcomes and cost-effectiveness. *Br J Psychiatry Suppl* 2010;53:s14–19.

8. Beck AT, Ward CH, Mendelson M, *et al*. An inventory for measuring depression. *Arch Gen Psychiatry* 1961;4:561–71.

9. Spielberger CD, Gorsuch RL, Lushene R, *et al*. Manual for the state-trait anxiety inventory. Palo Alto, CA: Consulting Psychologists Press, 1983.

10. Kay SR, Fiszbein A, Opler LA. The positive and negative syndrome scale (PANSS) for schizophrenia. *Schizophr Bull* 1987;13:261–76.

11. Trauer T. Outcome measurement in chronic mental illness. *Int Rev Psychiatry* 2010;22:99–113.

12. Wing JK, Beevor AS, Curtis RH, *et al*. Health of the Nation Outcome Scales (HoNOS). Research and development. *Br J Psychiatry* 1998;172:11–18.

13. Wing JK, Curtis R, Beevor A. *HoNOS: health of the nation outcome scales: report on research and development July 1993–December 1995*. London: Royal College of Psychiatrists, 1996.

14. Bebbington P, Brugha T, Hill T, *et al*. Validation of the Health of the Nation Outcome Scales. *Br J Psychiatry* 1999;174:389–94.

15. Sharma VK, Wilkinson G, Fear S. Health of the Nation Outcome Scales: a case study in general psychiatry. *Br J Psychiatry* 1999;174:395–8.

16. Trauer T, Callally T, Hantz P, *et al*. Health of the Nation Outcome Scales. Results of the Victorian field trial. *Br J Psychiatry* 1999;174:380–8.

17. Pirkis J, Burgess P, Coombs T, *et al*. Routine measurement of outcomes in Australia's public sector mental health services. *Aust New Zealand Health Policy* 2005;2:8.

18. Duke B. HoNOS in the consultation liaison psychiatry setting: is it valid? *Australas Psychiatry* 2010;18:547–50.

19. Orrell M, Yard P, Handysides J, *et al*. Validity and reliability of the Health of the Nation Outcome Scales in psychiatric patients in the community. *Br J Psychiatry* 1999;174:409–12.

20. Nunnally J, Bernstein I. *Psychometric theory*. 3rd edn. New York: McGraw-Hill, 1994.

21. Fleiss JL, Levin B, Paik MC. The measurement of interrater agreement. *Stat Methods Rates Proportions* 1981;2:212–36.

22. Shergill SS, Shankar KK, Seneviratna K, *et al*. The validity and reliability of the Health of the Nation Outcome Scales (HoNOS) in the elderly. *J Mental Health* 1999;8:511–21.

23. Strategic Planning Group for Private Psychiatric Services. A national model for the collection and analysis of a minimum data set with outcome measures for private, hospital-based, psychiatric services. Canberra: Commonwealth of Australia, 2000.

24. Mental Health and Drugs Division. Outcome measurement in clinical mental health services. Melbourne, Australia: Victorian Government Department of Human Services, 2007.

25. Barwon Health, Mental Health, Drugs and Alcohol Services. *ACE Guidelines: assessment, care and evaluation manual*. Barwon Health, 2010.

26. Parabiaghi A, Kortrijk HE, Mulder CL. Defining multiple criteria for meaningful outcome in routine outcome measurement using the Health of the Nation Outcome Scales. *Soc Psychiatry Psychiatr Epidemiol* 2014;49:291–305.

27. Egger ST, Weniger G, Prinz S, *et al*. Health of the Nation Outcome Scales in a psychiatric inpatient setting: assessing clinical change. *J Eval Clin Pract* 2015;21:236–41.

28. Speak BL, Hay P, Muncer SJ. HoNOS—their utility for payment by results in mental health. *Int J Health Care Qual Assur* 2015;28:115–28.

29. Lovaglio PG, Monzani E. Health of the nation outcome scales evaluation in a community setting population. *Qual Life Res* 2012;21:1643–53.