TECHNICAL REPORT:
Three Pilot Sites in Nova Scotia for the
Health Of the Nations Outcome Scales

Mental Health Monitoring Working Group

Crown copyright, Province of Nova Scotia, 2004
ISBN 0-88871-889-6

October, 2004
Executive Summary

**Aim:** To test the acceptability, usability, sensitivity and validity of a measure of mental health outcome, the Health of the Nation Outcomes Scales (HoNOS), under conditions that would resemble as closely as possible how it would be routinely used if it were introduced across Nova Scotia.

**Method:** Following training, 3 pilot sites provided ratings on all inpatient or outpatient referrals over 4 months using versions of HoNOS that cover children and adolescents (HoNOSCA), working age adults and the over-65s. Where possible, data were entered using the routine administrative data system of the Province. In the case of outpatients we used the Mental Health Outpatient Information System (MHOIS). Clinician acceptability was assessed by the percentage of patient contacts that were eligible for inclusion in the study and for which HoNOS ratings had been completed and by survey at the end of the project.

**Results:** Eight hundred initial patient contacts were eligible for rating with HoNOS and inclusion in the study, of whom just under 61% had at least one HoNOS rating (n=485). However, once patient contacts were rated and included in the study, subsequent completion rates were much higher especially among outpatients where 85% of eligible patient contacts had at least one follow-up rating. HoNOS ratings were also sensitive to time and setting, with significantly higher scores in inpatients than outpatients. The improvement over time was most evident for the global score but was also seen for some of the items such as Self-harm, Depression and Other emotional problems like anxiety. The differing pattern of scores between individual diagnoses lent further support to the validity of the scale. Although clinicians found the instrument and glossary easy to use, they remained undecided about the application of HoNOS to their clinical practice.

**Conclusions:** HoNOS appears to have satisfactory feasibility, acceptability, validity and sensitivity for routine use in Nova Scotia. More work is required to improve staff support for the routine collection of data on outcomes. Regular feedback of clinically useful information to those clinicians who collect the data can improve the quantity and quality of data. We hope to supplement the
paper-based reports to clinicians with access to a web-based database where staff can examine and
analyse their own patient data using simple descriptive statistics. We also recommend simplification
of the child and adolescent version with which staff had most difficulty.

**Implications**

We recommend that this pilot form the basis for the introduction of routine assessment of outcomes
across the Province using HoNOS. We suggest implementation through the immediate following steps:

- Deliver training and follow-up sessions for all clinicians to ensure reliability and fidelity of ratings
- Incorporate HoNOS ratings into MHOIS returns across the Province
- Simplify the child & adolescent form by provision of a tabular glossary that is similar to the format of the adult version and the dropping of two items that do not contribute to the total score
- Make resources available for the evaluation of the introduction of HoNOS after the first year of implementation

We would suggest that introduction of HoNOS into inpatient services be deferred until the development of routine data collection that mirrors MHOIS.

Ultimately, HoNOS should form part of the comprehensive measurement of patient outcome, also including quality of life, self-ratings, satisfaction with services, and family or carer views.
Introduction

Health service planners and providers need to measure patient outcomes to demonstrate that the mental health programmes are achieving intended results. However, measurement of mental health outcomes is problematic because they are often long-term, complex and multidimensional. Mental health outcomes are not only about symptoms, but also about subtle factors that are often difficult to operationalise, such as the ability to make and sustain relationships. It is therefore difficult to develop meaningful outcome measures, as the need to capture this complex information must be balanced against simplicity and ease of use of the instrument (1-2). Performance monitoring of mental health services has therefore relied on measures of structure and process that assess compliance with strategy or policy (3).

In response to this problem, the Nova Scotia Department of Health commissioned a pilot study on the introduction of a routine measure of clinical outcome across the Province. Such a measure needed to cover a wide range of symptoms, be easy to use and sensitive to changes in mental state. The Health of the Nation Outcomes Scales (HoNOS) is one such instrument. It has been validated in studies in Great Britain and Australia (4-10), although well-designed training and follow-up sessions are required to ensure reliability and fidelity of ratings (5). Unlike other instruments it covers all age groups, with versions for children and adolescents (HoNOSCA), working age adults and the over-65s (11-13). The scale is the same for working age adults and the over-65s, the difference between the two being in the accompanying glossaries (11, 12). We refer to these two forms of the scale in this paper as the adult version. Further versions for learning disability, forensic services and acquired brain injury are under development (14, 15). The HoNOS has 12 items (13 in the case of the version for children and adolescents) and four sections: behaviours, impairments, symptoms and social functioning (11-13). The adult version has an item for Other symptoms where symptoms other than delusions, hallucinations or depression are recorded (11, 12). Raters select the severest symptom in the previous 2 weeks from the following: phobia; anxiety; obsessive-compulsive disorder; mental strain/tension; dissociative symptoms; somatoform symptoms; eating; sleep; sexual; and other. HoNOSCA has two additional items on the lack of knowledge about the presenting difficulties or potential services (Section B), but these do not contribute to the total score (7, 13). Each scale is scored 0 to 4 with a total score of 48 (52 in the case of HoNOSCA). A higher
score indicates more impaired functioning. The HoNOS takes approximately 5 minutes per person to complete following a clinical interview.

The aim of this study was to test the acceptability, usability, sensitivity and validity of HoNOS in three pilot sites in Nova Scotia, under conditions that would resemble as closely as possible how it would be routinely used if it were introduced across the Province.

**Method**

**Setting**

The pilot study was carried out in three mental health services in Nova Scotia (South Shore, Colchester and Cole Harbour Health Services). All new inpatient or outpatient encounters were eligible for rating with the HoNOS. These HoNOS ratings were included in the study.

All staff had previously attended a workshop explaining the rationale for measuring outcomes in health services and potential benefits for clinicians and health service planners. Comments from staff informed the planning of the project. For instance, we had originally intended to only collect and give feedback to staff on global scores but broadened this to feedback on individual items as clinicians said this would be more useful to them.

**Training**

Staff at the three sites (n=68) were trained by one trainer (SK) either face-to-face or using telehealth. SK had previously attended a training course in the rating of HoNOS by the Royal College of Psychiatrists Research Unit in London. All staff received training in the adult versions (working age adults and over-65s). Mental health professionals who worked in child and adolescent services received a further half-day’s training based on a training video. We used written case vignettes to assess the intraclass correlation coefficients (ICC) at the end of training. The ICC for the overall global score was 0.96 (p=0.001). ICC scores for the some of the subsections were lower but still significant: Behaviour (0.99); Impairment (0.99); Symptoms (0.86); Social (0.92)
Collection of data

Data were collected from October 2003 to February 2004. All inpatient or outpatient referrals to each of the three sites from October to December were eligible for inclusion in the study. For inpatients, assessments were completed on admission and discharge for the duration of the pilot study. For outpatients, assessments were completed for each new referral from October to December 2003 and at each contact from October 2003 to February 2004. This was to ensure each subject had at least two assessments if possible.

Where possible, data were entered using the routine administrative data system of the Province. In the case of outpatients we used the Mental Health Outpatient Information System (MHOIS), which records demographics, ICD-9 diagnoses and care episodes for all outpatient contacts. Clinicians attached the HoNOS scores to the returns they submitted on each patient seen. The Mental Health Outpatient Information System generates Foxfire reports, which are submitted to the Department of Health. We also used these Foxfire reports to give participating clinicians feedback in the form of summary statistics on their patients' scores. We could not use the equivalent system for inpatients (the Discharge Abstract Database) as these data are collated in Ottawa for the whole of Canada before being returned to the site of origin. This would have meant a three-month delay before we were able to analyse the data. Inpatient data were therefore entered on an Access database and also submitted to the Department of Health. All outpatient and inpatient data were then dumped into an Excel file and analysed using the Statistical Package for Social Sciences (SPSS).

We assessed clinician acceptability of the use of the instrument by two methods: the percentage of patient contacts which were eligible for inclusion in the study for which HoNOS ratings had been completed, as well as five-point Likert scales that were completed by participating clinicians at the end of the pilot study.

Statistical analysis

We assessed concurrent validity by measuring the association between mean HoNOS scores and
both diagnosis and setting (inpatient v community) using the Students’ T-test or analysis of variance as appropriate. We assessed change over time using the paired samples T-test for both the global scores and each item. We also compared differences in mean changes in HoNOS score between versions (child and adolescent, or adult forms) and setting (inpatients or outpatients). For each item we also assessed change from the presence of symptomatology to none or minimal symptomatology using Wilcoxon’s Sign Rank tests. We assessed the Likert scales of clinicians’ views in the same way.

Results

The data

Eight hundred initial patient contacts were eligible for inclusion in the study, for which just under 61% had at least one HoNOS rating (n=485). Only thirty eight percent of eligible consultations with a psychiatrist resulted in a completed HoNOS rating compared to 68% of consultations with other occupational groups (nurses, occupational therapists, psychologists or social workers), this result being highly statistically significant (chi square=53.8, df=1, p<0.0001).

Of the 485 subjects who had had at least one HoNOS rating, 277 (57%) were female. The average age of the sample was 26.8 (SD=17.4), although data on age were missing on thirty subjects. Fifty subjects were admitted at least once to hospital during the pilot study. Eighty five percent were rated using the adult version (n=412), the remaining 73 using the child and adolescent version.

Concurrent validity

The mean HoNOS score was 17.1 for inpatients (SD=5.38) compared to 10.8 for outpatients (SD=5.85), this result reaching statistical significance (t-test=1.97, df=466, p=0.04). There was no statistical difference between mean scores for the adult version (10.6, SD=5.7) and HoNOSCA (11.1, SD=5.9).

We analysed the adult version and HoNOSCA separately in the following analyses. Using the
methodology of Trauer et al and Lauzon et al (6, 16), we compared the mean scores of each item by sex, and three age groups (less than 39 years, 40 –49 years, 50 years and over). Males had significantly higher ratings on Alcohol/drugs (mean=0.67, SD=0.8) than females (mean=0.4, SD=0.8) ($t=-2.52$, $df=410$, $p=0.01$). By contrast, females had higher ratings on Depression (mean=1.57, SD=1.1) than males (1.27, SD=1.0) ($t=2.39$, $df=410$, $p=0.02$). In comparison with the other two groups, patients over 50 years were significantly less likely to use alcohol or drugs (means (SD) of 0.19 (0.6), 0.6 (1.0) and 0.63 (1.0)) respectively ($F=2.31$, $p=0.001$), but more likely to have higher scores for Cognitive impairments (means (SD) of 0.68 (0.9), 0.46 (0.7) and 0.44 (0.7)) ($F=3.28$, $p=0.04$) as well as Physical problems (means (SD) of 1.20 (1.2), 0.67 (1.0) and 0.64 (1.0)) ($F=9.06$, $p<0.0001$).

Table 1 shows the mean HoNOS scores by main diagnostic groups for the adult version. Organic psychiatric disorders had the highest total score, with particularly high scores in Cognitive and Physical impairment. Patients with non-affective and affective psychoses had similar patterns other than an increased score for depressive symptoms for the latter. Patients in the depressive group had highest ratings on Depression and Other symptoms. As might be expected, patients with substance or alcohol use had high ratings on the Alcohol /drug item. Patients with anxiety had the highest scores for Other symptoms where anxiety symptoms are recorded. Patients with personality disorders had very high scores, second only to organic disorders, and higher than the score for psychotic disorders. They had particularly high scores in Other symptoms and Relationships.
Table 1: Mean HoNOS ratings by main diagnostic groups (Adult version)

<table>
<thead>
<tr>
<th>Diagnosis</th>
<th>N</th>
<th>Aggression</th>
<th>Self-harm</th>
<th>Alcohol/drug</th>
<th>Cognition</th>
<th>Physical</th>
<th>Hallucinations delusions</th>
<th>Depression</th>
<th>Other symptoms</th>
<th>Relationships</th>
<th>Activities of Daily Living</th>
<th>Accommodation</th>
<th>Occupation</th>
<th>Total Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Organic disorders</td>
<td>4</td>
<td>1.75</td>
<td>0.00</td>
<td>0.00</td>
<td>2.25</td>
<td>1.75</td>
<td>1.03</td>
<td>1.25</td>
<td>1.50</td>
<td>1.50</td>
<td>1.25</td>
<td>0.50</td>
<td>0.75</td>
<td>13.53</td>
</tr>
<tr>
<td>Substance/ETOH use</td>
<td>6</td>
<td>0.50</td>
<td>0.33</td>
<td>2.00</td>
<td>0.00</td>
<td>1.00</td>
<td>0.00</td>
<td>2.17</td>
<td>2.00</td>
<td>2.50</td>
<td>0.20</td>
<td>0.50</td>
<td>1.50</td>
<td>13.00</td>
</tr>
<tr>
<td>Non-affective psychosis</td>
<td>84</td>
<td>0.39</td>
<td>0.15</td>
<td>0.34</td>
<td>0.58</td>
<td>0.64</td>
<td>1.18</td>
<td>0.76</td>
<td>1.29</td>
<td>1.58</td>
<td>1.06</td>
<td>0.56</td>
<td>1.38</td>
<td>9.75</td>
</tr>
<tr>
<td>Affective psychosis</td>
<td>22</td>
<td>0.50</td>
<td>0.19</td>
<td>0.81</td>
<td>0.56</td>
<td>0.56</td>
<td>0.00</td>
<td>1.44</td>
<td>1.69</td>
<td>1.31</td>
<td>1.00</td>
<td>0.50</td>
<td>0.88</td>
<td>9.44</td>
</tr>
<tr>
<td>Depression</td>
<td>73</td>
<td>0.50</td>
<td>0.86</td>
<td>0.65</td>
<td>0.50</td>
<td>0.95</td>
<td>0.18</td>
<td>2.23</td>
<td>2.37</td>
<td>1.77</td>
<td>1.05</td>
<td>0.45</td>
<td>1.23</td>
<td>12.75</td>
</tr>
<tr>
<td>Anxiety</td>
<td>38</td>
<td>0.41</td>
<td>0.32</td>
<td>0.18</td>
<td>0.41</td>
<td>0.74</td>
<td>0.15</td>
<td>1.53</td>
<td>2.71</td>
<td>1.18</td>
<td>0.62</td>
<td>0.29</td>
<td>1.18</td>
<td>9.58</td>
</tr>
<tr>
<td>Adjustment disorder</td>
<td>69</td>
<td>0.52</td>
<td>0.31</td>
<td>0.51</td>
<td>0.39</td>
<td>1.15</td>
<td>0.15</td>
<td>1.53</td>
<td>1.98</td>
<td>1.76</td>
<td>0.84</td>
<td>0.40</td>
<td>1.05</td>
<td>10.39</td>
</tr>
<tr>
<td>Personality disorder</td>
<td>10</td>
<td>1.80</td>
<td>0.60</td>
<td>1.00</td>
<td>0.50</td>
<td>1.10</td>
<td>0.10</td>
<td>1.20</td>
<td>2.40</td>
<td>2.80</td>
<td>0.70</td>
<td>0.70</td>
<td>0.90</td>
<td>13.44</td>
</tr>
<tr>
<td>Childhood disorder</td>
<td>7</td>
<td>0.86</td>
<td>0.00</td>
<td>0.29</td>
<td>1.00</td>
<td>0.43</td>
<td>0.14</td>
<td>0.86</td>
<td>1.86</td>
<td>1.29</td>
<td>0.86</td>
<td>0.43</td>
<td>1.14</td>
<td>9.14</td>
</tr>
<tr>
<td>Other/ non-specific</td>
<td>99</td>
<td>0.86</td>
<td>0.45</td>
<td>0.49</td>
<td>0.50</td>
<td>0.55</td>
<td>0.28</td>
<td>1.32</td>
<td>1.55</td>
<td>1.46</td>
<td>0.61</td>
<td>0.40</td>
<td>0.82</td>
<td>9.16</td>
</tr>
<tr>
<td>F</td>
<td>3.77</td>
<td>2.83</td>
<td>2.93</td>
<td>3.24</td>
<td>2.02</td>
<td>12.69</td>
<td>8.09</td>
<td>5.48</td>
<td>2.47</td>
<td>1.69</td>
<td>0.53</td>
<td>1.23</td>
<td>2.39</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>412</td>
<td>0.61</td>
<td>0.41</td>
<td>0.51</td>
<td>0.51</td>
<td>0.80</td>
<td>0.35</td>
<td>1.46</td>
<td>1.89</td>
<td>1.62</td>
<td>0.85</td>
<td>0.45</td>
<td>1.10</td>
<td>10.45</td>
</tr>
</tbody>
</table>

* administered the adult version not HoNOSCA *** p<0.0001 ** p<0.01 * p<0.05

We assessed whether these differences in the symptom profile between disorders reached statistical significance using analysis of variance. There were significant differences between diagnoses for all items except activities of daily living (ADL), Accommodation and Occupation (Table 1).

There were 265 ratings of subsidiary codes in item 8, Other symptoms. Anxiety was the most common condition (35%), followed by strain/tension (30%) and sleep disturbances (18%). All the
other conditions were comparatively rare and none exceeded 5%.

For HoNOSCA (Table 2), the only difference between the sexes was for a significantly higher score for Emotional problems including anxiety and depression in females (mean (SD)=2.56 (1.0)) than males (mean (SD)=1.83 (1.2) (t-test=2.54, df=71, p=0.01)). As expected, there was an association between age and both Self-harm and Alcohol/drug use. Patients over 12 years were significantly more likely than younger patients to have self-harmed (means (SD) of 0.65 (1.2) and 0.09 (0.3) respectively) (t-test=-2.59, df=71, p=0.01), or used alcohol and drugs (means (SD) of 0.59 (1.0) and 0.03 (0.1) respectively) (t-test=-3.21, df=71, p=0.002).

Table 2: Mean HoNOS ratings by main diagnostic groups (Children & adolescents)

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>Aggression</th>
<th>Over-activity</th>
<th>Self-harm***</th>
<th>Alcohol/Drug***</th>
<th>Scholastic</th>
<th>Physical</th>
<th>Hallucinations/ Delusions</th>
<th>Non-organic somatic</th>
<th>Emotional Problems**</th>
<th>Peers</th>
<th>Self care</th>
<th>Family</th>
<th>School attendance</th>
<th>Total score *</th>
</tr>
</thead>
<tbody>
<tr>
<td>Depression</td>
<td>10</td>
<td>1.22</td>
<td>1.33</td>
<td>1.78</td>
<td>1.22</td>
<td>1.67</td>
<td>0.56</td>
<td>0.11</td>
<td>1.33</td>
<td>1.56</td>
<td>0.56</td>
<td>2.33</td>
<td>1.22</td>
<td>16.78</td>
<td></td>
</tr>
<tr>
<td>Anxiety</td>
<td>6</td>
<td>1.33</td>
<td>1.33</td>
<td>0.00</td>
<td>0.17</td>
<td>0.50</td>
<td>0.00</td>
<td>0.00</td>
<td>0.67</td>
<td>3.00</td>
<td>2.00</td>
<td>0.33</td>
<td>0.83</td>
<td>10.17</td>
<td></td>
</tr>
<tr>
<td>Adjustment disorder</td>
<td>13</td>
<td>1.38</td>
<td>1.46</td>
<td>0.46</td>
<td>0.08</td>
<td>0.75</td>
<td>0.23</td>
<td>0.46</td>
<td>0.92</td>
<td>2.46</td>
<td>1.23</td>
<td>0.54</td>
<td>1.92</td>
<td>11.08</td>
<td></td>
</tr>
<tr>
<td>Childhood disorder</td>
<td>20</td>
<td>1.42</td>
<td>1.84</td>
<td>0.16</td>
<td>0.21</td>
<td>1.42</td>
<td>0.16</td>
<td>0.00</td>
<td>0.47</td>
<td>1.61</td>
<td>1.32</td>
<td>0.84</td>
<td>1.63</td>
<td>10.56</td>
<td></td>
</tr>
<tr>
<td>Other/ non-specific</td>
<td>24</td>
<td>1.13</td>
<td>1.08</td>
<td>0.08</td>
<td>0.25</td>
<td>0.96</td>
<td>0.33</td>
<td>0.38</td>
<td>0.58</td>
<td>1.71</td>
<td>0.96</td>
<td>0.71</td>
<td>1.71</td>
<td>9.88</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>73</td>
<td>1.28</td>
<td>1.41</td>
<td>0.38</td>
<td>0.32</td>
<td>1.10</td>
<td>0.27</td>
<td>0.23</td>
<td>0.72</td>
<td>2.11</td>
<td>1.27</td>
<td>0.66</td>
<td>1.73</td>
<td>11.19</td>
<td></td>
</tr>
</tbody>
</table>

*** p<0.0001  ** p=0.01  * p=0.05

Table 2 shows the mean scores by main diagnostic groups for HoNOSCA. Patients with depression had the highest Self-harm, Alcohol/drug use and Total scores, the difference between diagnoses reaching statistical significance. There were no other statistical differences between items. Scores
for delusions and hallucinations were low in all diagnoses.

**Change over time**

The instrument's sensitivity to change was assessed by rating patients on two or more occasions up to 4 months apart. For inpatients, this occurred at admission and discharge, for outpatients at every contact. Of 50 inpatients, six were still in hospital at the end of the pilot leaving 44 who should have had a rating done at discharge. Of these 30 (68%) had had a second rating done. For outpatients, 363 were eligible for follow-up, of whom 310 (85%) had at least one follow-up rating. The interval between ratings was an average of 24.3 days (SD=28.5).

Two hundred and fifteen outpatients had a third rating during the pilot study, at an average of 45.5 days (SD=29.4) after their initial assessment. One hundred and thirty five outpatients had a fourth assessment at an average of 58.6 days (SD=27.6) after entry into the pilot study.

Table 3 shows the mean total scores at each rating. There were significant decreases in HoNOS score over the length of the project for both inpatients and outpatients.

**Table 3: Results of follow-up**

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>Mean</th>
<th>SD</th>
<th>Comparison groups</th>
<th>t-test</th>
<th>df</th>
<th>Signif.</th>
</tr>
</thead>
<tbody>
<tr>
<td>IP discharge total HoNOS score</td>
<td>30</td>
<td>9.33</td>
<td>5.71</td>
<td>IP admission total HoNOS score - IP discharge total HoNOS score</td>
<td>8.05</td>
<td>29</td>
<td>0.000</td>
</tr>
<tr>
<td>OP total score 2</td>
<td>310</td>
<td>10.34</td>
<td>5.71</td>
<td>OP total score 1 - OP total score</td>
<td>3.12</td>
<td>309</td>
<td>0.002</td>
</tr>
<tr>
<td>OP total score 3</td>
<td>215</td>
<td>9.41</td>
<td>5.27</td>
<td>OP total score 1 - OP total score</td>
<td>4.34</td>
<td>214</td>
<td>0.000</td>
</tr>
<tr>
<td>OP total score 4</td>
<td>135</td>
<td>9.80</td>
<td>5.28</td>
<td>OP total score 1 - OP total score</td>
<td>4.44</td>
<td>134</td>
<td>0.000</td>
</tr>
</tbody>
</table>
Figure 1 shows the reduction in scores for those outpatients who had had 4 ratings during the course of the pilot study (n=135).

**Figure 1: Reduction in scores for outpatients with 4 ratings**

We were only able to analyse the change in individual items for outpatients as our inpatient sample was too small for analysis. We analysed the change in individual items from the first to the third rating for the adult version (n=183) and HoNOSCA (n=32) separately (Table 4). We used the third rating to ensure comparability with previous work that had reported the results at between four and six week follow-up (5, 6). For the adult version, there were significant changes in items measuring Aggression, Self-harm, Depression and Other symptoms. For HoNOSCA, there were significant changes in items measuring Aggression, Overactivity, Alcohol/drug use, Non-organic somatic symptoms, Emotional symptoms and School Attendance.
We found similar results when we assessed change from the presence of symptomatology to none or minimal symptomatology for each item (Table 4)

**Table 4: Changes in individual HoNOS items for outpatients**

<table>
<thead>
<tr>
<th>Pair</th>
<th>H_n  - H2_n</th>
<th>Continuous variables</th>
<th>Change from 2/4 to 0/1</th>
<th>Continuous variables</th>
<th>Change from 2/4 to 0/1</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>t (df=182)</td>
<td>Sig. (2-tailed)</td>
<td>Z</td>
<td>Sig. (2-tailed)</td>
</tr>
<tr>
<td>Pair 1</td>
<td>H_1 - H2_1</td>
<td>1.18</td>
<td>0.24</td>
<td>-0.73</td>
<td>0.47</td>
</tr>
<tr>
<td>Pair 2</td>
<td>H_2 - H2_2</td>
<td>2.88</td>
<td>0.00</td>
<td>-3.13</td>
<td>0.00</td>
</tr>
<tr>
<td>Pair 3</td>
<td>H_3 - H2_3</td>
<td>1.59</td>
<td>0.11</td>
<td>-2.00</td>
<td>0.05</td>
</tr>
<tr>
<td>Pair 4</td>
<td>H_4 - H2_4</td>
<td>-0.22</td>
<td>0.83</td>
<td>-0.43</td>
<td>0.67</td>
</tr>
<tr>
<td>Pair 5</td>
<td>H_5 - H2_5</td>
<td>-0.85</td>
<td>0.40</td>
<td>-0.47</td>
<td>0.64</td>
</tr>
<tr>
<td>Pair 6</td>
<td>H_6 - H2_6</td>
<td>1.31</td>
<td>0.19</td>
<td>-1.70</td>
<td>0.09</td>
</tr>
<tr>
<td>Pair 7</td>
<td>H_7 - H2_7</td>
<td>2.31</td>
<td>0.02</td>
<td>-2.56</td>
<td>0.01</td>
</tr>
<tr>
<td>Pair 8</td>
<td>H_8 - H2_8</td>
<td>2.29</td>
<td>0.02</td>
<td>-7.31</td>
<td>0.00</td>
</tr>
<tr>
<td>Pair 9</td>
<td>H_9 - H2_9</td>
<td>2.19</td>
<td>0.03</td>
<td>-1.37</td>
<td>0.17</td>
</tr>
<tr>
<td>Pair 10</td>
<td>H_10 - H2_10</td>
<td>1.55</td>
<td>0.12</td>
<td>-1.37</td>
<td>1.00</td>
</tr>
<tr>
<td>Pair 11</td>
<td>H_11 - H2_11</td>
<td>0.14</td>
<td>0.89</td>
<td>-0.77</td>
<td>0.44</td>
</tr>
<tr>
<td>Pair 12</td>
<td>H_12 - H2_12</td>
<td>1.53</td>
<td>0.13</td>
<td>-0.87</td>
<td>0.38</td>
</tr>
<tr>
<td>Pair 13</td>
<td>H_13 - H2_13</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

H_n = Item rated at baseline   H2_n = Item rated at follow-up (mean of 45 days after baseline)

**Clinicians’ viewpoints**

Thirty-six clinicians from the three pilot sites returned assessments of their participation in the pilot project. The mean time taken to complete the HoNOS was 5 minutes. We also asked clinicians about their experience of using the instrument by means of five-point Likert scales (0=strongly disagree, 1=disagree, 2=neutral, 3=agree, 4=strongly agree). We present medians given the ordinal
nature of the scale and non-normal distribution of the data. In terms of the process of using HoNOS, median scores for ease of use and usefulness of the glossary indicate they were easy to use (Figures 2 and 3).

Using similar scales (0=strongly disagree, 1=disagree, 2=neutral, 3=agree, 4=strongly agree), clinicians were also asked if there were any individual items that were difficult to score. Some items were difficult to score (median score=3), with item 8 from the adult version being identified
in 80% of cases. In the same fashion, clinicians were asked about the usefulness of HoNOS ratings in their clinical practice. Clinicians were undecided about the usefulness of the feedback, the relevance of the information, the sensitivity to change, the utility of the social items or help with planning services from using HoNOS (median scores for all items=2) (Figure 4). We analysed any differences between those clinicians who had used the adult version (n=25) and those using HoNOSCA (n=11). Clinicians found the HoNOSCA glossary significantly less useful in comparison with the adult version (medians of 2 v 3, z=-2.52, p=0.01). They also found HoNOSCA proved less useful in terms of relevant information (medians of 1 v 3, z=-3.58, p<0.0001), sensitivity to change (medians of 0 v 3, z=-2.49, p=0.02), or social items (medians of 0 v 3, z=-2.17, p=0.03)

![Figure 4: Perceived sensitivity to change](image)

**Discussion**

*Strengths*

This is the first study to evaluate the introduction of HoNOS across all age groups into routine clinical care in a pilot that resembled normal clinical practice as much as possible. Clinicians collected all the ratings in the course of routine clinical work. No research workers were involved. All relevant staff had attended training and participated in an assessment of inter-rater reliability, although we relied on a written case vignette that may have elevated our ICC. Importantly, 90% of
the data were captured using existing routine administrative collection. We included clinicians’ comments in the planning of the pilot project and provided them with feedback on their own patients. This is also the first study of HoNOS in North America in an English-speaking population, the only previous publication being a validation of a French version in Quebec (16). Wherever possible, we have used methodologies that allow comparisons with other jurisdictions.

It is unlikely that the changes in HoNOS scores we reported could have been influenced by previous HoNOS ratings. Scores were submitted to the Department of Health after each consultation and did not remain in the clinical notes.

**Weaknesses**

In spite of attempts to consult and involve clinicians throughout the project, we achieved a relatively low return rate of 61% that may affect the generalisability of our findings. One explanation might be that although clinicians found the instrument and glossary easy to use, they remained undecided about the application of HoNOS to their clinical practice. However, once initial ratings were included in the study, completion rates were much higher, especially among outpatients where 85% of eligible patients had at least one follow-up rating. Also, we assessed patients over a relatively short time and it is possible that some changes such as those in social problems may take longer to become apparent. However, it is interesting to note that the biggest change in global scores occurred in the first 45 days of follow-up. Because of the differences in some of the items between the adult version and HoNOSCA, we had to analyse these separately when looking at some of the items. This further reduced study power, although statistically significant results did emerge even within the relatively small group who were assessed using HoNOSCA. Lastly, our reliance on routine administrative datasets to collect our data meant we had information on a limited number of socio-demographic factors, and only on ICD-9 diagnoses.

**Validity and sensitivity**

In terms of the cross-sectional findings, our global scores for inpatients and outpatients are strikingly similar to those reported in Great Britain, Australia and French Canada (4-9, 16). Our
HoNOS ratings were also sensitive time and setting, with significantly higher scores in inpatients than outpatients. The improvement over time was most evident for the global score but was also seen for some of the items such as Self-harm, Depression and Other emotional problems like anxiety. In HoNOSCA, there were also significant changes in Aggression and School attendance items. By contrast, items measuring Psychosis, Physical problems, Cognitive impairment and Social issues showed little change. Some of this may represent the longer-term nature of these problems and the need for longer-term follow-up to detect improvement. HoNOSCA seemed particularly sensitive given the small number of patients who had follow-up scores. This was in spite of it being rated by clinicians as significantly less useful in terms of sensitivity and relevance.

We also assessed the change in items from the presence of symptomatology to none or minimal symptomatology and found similar results. We made this comparison because clinicians stated that reporting change in this way is more useful and clinically relevant.

These findings are similar to results from Britain where change in HoNOSCA scores was greatest in items measuring symptoms and behaviours, rather than social and impairment scales (7). In the case of the adult version, findings are more equivocal with the suggestion that sensitivity of ratings is associated with the degree of training received by raters (5, 9).

In terms of differences between diagnoses, patients with organic disorders had the highest global HoNOS score. However we also found, as in work from Australia, that non-psychotic conditions such as depression and personality disorder had higher scores than the psychoses (6). In the case of personality disorder, one explanation for increased scores has been the presence of comorbid substance use that may not be captured when relying only on primary psychiatric diagnoses (6). Certainly, patients who had a primary diagnosis of alcohol or substance use also had high global scores in our study.

The differing pattern of scores between individual diagnoses lends further support to the validity of the scale and is also similar to work from other jurisdictions (5, 16). Men had higher rating for drug and alcohol use than females, while those who were over 50 scored more highly on problems with cognition and physical health. In line with previous research, children over 12 had significantly
higher scores than younger patients for HoNOSCA items on suicidality and alcohol or drug use (7).

**Implications**

HoNOS appears to have satisfactory feasibility, acceptability, validity and sensitivity for routine use in an English-speaking Canadian setting. It is relatively simple to explain and use. The scores provide a record of clinical progress and a tool for clinical audit and research. It is possible to introduce the adult version and HoNOSCA simultaneously and collect the data using routine databases. All the versions appear to perform in a similar way to studies in other jurisdictions.

We hope that this pilot will form the basis for the introduction of routine assessment of outcomes across Nova Scotia. The linked administrative databases in the Province give the potential for patient-based records covering inpatient, outpatient and community and primary care settings that could track clinical progress across all health care settings. As nearly 30,000 patients a year are seen by mental health services with an average number of 5.5 visits, this would give sufficient power to study the association of HoNOS scores with health service use including length of stay, admission rates, ambulatory and primary care contacts. These numbers would allow the use of logistic and multiple regression to control for covariates and confounders such as demographic variables, setting and diagnosis (17). A study from Australia suggested that bed use showed a good correlation with four HoNOS items: Aggression, Cognitive impairment, Hallucinations and delusions, and impaired Activity of daily living, but this study has not been replicated elsewhere (6, 18). If the same were to hold true of Nova Scotia, this would help quantify the need for hospital beds.

An initial step would be the incorporation of HoNOS ratings in MHOIS returns across the Province. The incorporation of HoNOS ratings into MHOIS has implications for staff development as well-designed training and follow-up sessions are required to ensure reliability and fidelity of ratings (5). We intend to use training videos, telehealth and training the trainer sessions to cover all mental health service staff in the Province.
More work is also required to improve staff support for the routine collection of data on outcomes. Regular feedback of clinically useful information to those clinicians who collect the data can improve the quantity and quality of data. We hope to supplement the paper-based reports to clinicians with access to a web-based database where staff can examine and analyse their own patient data using simple descriptive statistics. The significantly lower acceptance of HoNOSCA compared to the adult version merits further study, especially as the scales performed well in spite of the small numbers. One immediate change would be to convert the HoNOSCA glossary to the tabular form of the adult version, which was found to be significantly more useful. We also plan to omit the Section B ratings, which do not contribute to the total score (7), to shorten the number to 13 items.

We do not share the rather pessimistic view that clinicians necessarily see the completion of rating scales as an intrusion into the clinician-patient relationship that will not be tolerated for any length of time (18). Rather, we hope that HoNOS will form part of the comprehensive measurement of patient outcomes including quality of life, self-ratings, satisfaction with services, and family or carer views (6).

References
3. McEwan K, Goldner EM. Accountability and Performance Indicators for Mental Health


10. Trauer T. Use of a routine outcome measure in a consultation-liaison mental health service. Australasian Psychiatry, 2004, 12, 139-144


Contributors

Author: Dr. Stephen Kisely, Department of Psychiatry, Dalhousie University
Mental Health Monitoring Working Group

Doug Crossman, Co-Chair
John Campbell, Co-Chair
Bea MacConnell
Linda Judge
Steve Gleich
Karen Briand
Adrain MacKenzie
Susan Mercer
Charles Gilbert
Tony Prime, Project Officer
Susan Shaddick
Stephen Kisely