Reliable and Clinically Significant Change in Outcomes for Forensic Mental Health Inpatients: Use of the HoNOS-Secure

Geoffrey L. Dickens and Laura E. O'Shea

Reliable and clinically significant change in outcomes for forensic mental health inpatients: Use of the HoNOS-Secure

Geoffrey L. Dickens1 *

Laura E. O'Shea2

1 School of Social and Health Sciences, Abertay University, Bell Street Dundee and NHS Fife

2St Andrew's Academic Centre, Billing Road, Northampton and Abertay University, Bell Street, Dundee

*Corresponding author: Division of Mental Health Nursing and Counselling, Abertay University, Bell Street, Dundee. United Kingdom. DD1 1HG.

Email: g.dickens@abertay.ac.uk

Abstract

Secure mental health service outcomes are commonly derived from post-discharge recidivism, readmission and mortality rates. Information about less rare behavioural, functional, and symptom-related outcomes across the sometimes lengthy span of admission is scant. We aimed to determine whether patients underwent reliable and clinically significant change in risk-related need, behaviour, functioning, and symptoms during admission from analysis of routinely collected HoNOS-Secure data. Outcomes (*N*=418) were examined to determine between-group differences in baseline scores and rates of change, the proportion of patients for whom change was reliable, and, of those, the proportion whose scores fell during treatment by a clinically significant margin. Reliable change was demonstrated for just 4.8% and 5.7% of patients on the HoNOS-Secure scale and security scale respectively, and was rarely clinically significant. In a context in which services rarely report routinely collected data for a range of outcomes, this study found that HoNOS-Secure captured little of any clinical change that may have occurred. Further research is needed to determine whether the HoNOS-Secure is the most suitable tool for routine outcomes reporting, but, in any event, patients, clinicians, and the public should reasonably expect services to routinely demonstrate their effectiveness.

Key words: Outcomes, HoNOS-Secure, forensic mental health services, reliable clinical change

Introduction

The core functions of secure forensic mental health services include treatment of mental disorder, management of behavioural disturbance, and reducing the risk of harm to others (Department of Health, 2007). Multiple outcomes measures have been utilised in research studies in secure services (Chambers et al., 2009), but data which aims to measure the whole-service effect, as opposed to discrete treatment interventions, focuses almost exclusively on recidivism (Baxter, Rabe-Hesketh, & Parrott,1999; Buchanan, 1998; Coid, Hickey, Kahtan, Zhang, & Yang, 2007; Edwards, Steed, & Murray, 2002; Maden, Scott, Burnett, Lewis, & Skapinakis, 2004; Maden, Skapinakis, Lewis, Scott, Burnett, & Jamieson, 2006; Sahota, Davies, Duggan, & Clarke, 2009), readmission (Baxter et al., 1999; Davies, Clark, Hollin, & Duggan, 2007), and mortality (Baxter et al., 1999; Davies et al., 2007). There are very few data on changes in symptomatology, functioning, quality of life, recovery and perceived risk. Further, with rare exceptions (Edwards et al., 2002; O'Shea & Dickens, 2015), studies have reported outcomes solely after discharge from services and not those achieved during an inpatient admission. While data on violent recidivism and psychiatric relapse is essential, service providers and commissioners are increasingly concerned with service effectiveness in the in-patient setting, and for a wider range of therapeutic targets. Very nearly half (47%) of forensic inpatients are detained for over five years (Rutherford & Duggan, 2007), and it is incumbent upon service providers and commissioners to demonstrate service effectiveness during this time. One method of generating relevant data is through the use of routinely administered outcomes measures. In a systematic review of outcome measures used in forensic mental health services (Fitzpatrick et al., 2010) the most commonly used indicator in 308 (72%) separate studies were various measures of repeat offending behaviour. While prevention of recidivism is the ultimate goal of forensic services, the most severe and concerning offending is rare; this may hamper the utility of offending-related data

as outcomes measures since it requires long follow-up periods and large samples. While less

serious offending and risk assessment measures might also be used to indicate outcome, some

argue that the relative neglect of the clinical, rehabilitation and humanitarian domains

including quality of life, social functioning and psychosocial adjustment have at best been

underemployed (Fitzpatrick et al., 2010) and, at worst, have encouraged an overemphasis on

public safety over patient wellbeing (Cohen & Eastman, 2000).

It is more than a quarter of a century since US physician Paul Ellwood (1988) first

challenged health service providers to routinely measure and publish patients' aggregated

healthcare outcomes. This information would, he claimed, provide a transparent source of

information for service users and commissioners. Further, the data would complement that

from formal controlled trials of treatment efficacy, which are generally conducted with

patients who are highly selected to lack the co-morbidity found in reality, and thus provide

information about service effectiveness. Routine outcomes measurement, therefore, was

proposed to have the potential to provide a source of ecologically valid information to aid

evaluation of the complex social interventions which, alongside pharmacological and

psychological therapies, are a characteristic feature of inpatient mental health services

(Holloway, 2002). In the United Kingdom (UK), Australia, and New Zealand, the Health of

the Nation Outcome Scales (HoNOS; Wing, Beevor, Curtis, Park, Hadden, & Burns, 1998)

was selected as the tool of choice for measuring mental health service outcomes in the late

1990s. The HoNOS comprises 12 clinician-rated items relating to behaviour, impairment,

symptoms, and social functioning outcomes. The original HoNOS was designed for adults of

working age and was soon joined by a range of similar, but amended, tools for use with

specialist populations. One of these, the HoNOS-Secure (originally the HoNOS-MDO

[Mentally Disordered Offenders]), was developed for forensic mental health settings

(Dickens, Sugarman, & Walker, 2007; Sugarman & Walker; 2007) because parts of the

original HoNOS can be hard to interpret in secure settings, and the original scales did not capture a broad range of outcomes related to physical, procedural and relational security-need.

In the mental health and forensic mental health contexts, Ellwood's grand vision has had limited impact. In the UK, there have reportedly been problems with the HoNOS' data quality despite mandated reporting (Delaffon, Anwar, Noushad, Ahmed, & Brugha, 2012). In Australia, however, outcomes measurement with the HoNOS has been more robust (Pirkis, Burgess, Coombs, Clarke, Jones-Ellis, & Dickson, 2005). Burgess, Pirkis, and Coombs (2006) reported on routinely collated data from 38,351 inpatient and community acute mental health patients. The HoNOS scores improved over time, and were mediated by treatment setting and episode type. The HoNOS have also been used as a research outcome measure with mixed results. In New Zealand, Turner, Boden, Smith-Hamel, and Mulder's (2009) study of 236 early intervention psychosis service users reported improved HoNOS scores that the authors interpreted to reflect functional recovery over time. Improvements were mirrored on the Quality of Life Scale (Burckhardt & Anderson, 2003), and Global Assessment of Functioning (GAF; Hall, 1995) score. Conversely, in Scotland, a study of 789 people with schizophrenia spectrum disorders (Hunter, Cameron & Norrie, 2009) detected no significant change on the HoNOS behaviour or symptom subscales, improvement on the social functioning subscale, and decline on the impairment subscale. In contrast, statistically significant clinical improvement was detected on all subscales of the Avon Mental Health Measure (Markovitz, 1996) suggesting that the HoNOS was insufficiently sensitive to meaningful change in this group. Audin, Margison, Clark, and Barkham (2001). reported that there was little change on the majority of HoNOS items in 362 outpatients in a psychotherapy service and argued that the tool was unsuitable for measuring outcomes in that population. A small number of studies have examined change over time in HoNOS-Secure ratings made

during routine practice in a forensic mental health service (Dickens, Picchioni, Long, & Sugarman, 2010; Long, Dickens, Sugarman, Craig, Mochty, & Hollin, 2010; Sugarman & Dickens, 2010; Sugarman, Walker, & Dickens, 2009) and while there were small, statistically significant improvements in item and total scores none of these studies have examined whether change was either reliable, that is not simply due to measurement error, or clinically significant; nor whether any measured change differed between selected clinical and demographic groups. We have therefore conducted this retrospective, observational study using iterated clinical team ratings of individual patients' HoNOS-Secure assessments in order to determine whether scores change reliably and significantly over time, and whether there are between group differences in rates of change. The overall aim is to inform forensic mental health policymakers and clinicians about the utility of HoNOS-Secure as a meaningful routine outcomes measure.

Method

Setting and Participants

St Andrew's provides specialist secure mental health care across four sites in England. Accommodation is provided in gender-specific medium and low secure wards. Eligible patients were those adults who had at least two HoNOS Secure assessments completed, were resident in the mental health pathway, and did not have an intellectual disability.

Design and Procedure

This study employed a retrospective design; HoNOS-Secure assessments were completed by clinical teams as part of routine clinical practice between 2004 and 2014. In September 2014 we extracted a maximum of six consecutive assessments per patient in an anonymised format, along with demographic and clinical information. Since data was routinely collected, and was provided to the researchers in an anonymised format, the study

did not require approval by an NHS Research Ethics Committee. The study was approved by the clinical audit manager.

Measures

HoNOS Secure. The HoNOS-Secure version 2b comprises modified versions of the original twelve HoNOS items (Wing et al., 1998) which are intended to assess behaviour, impairment, symptoms, and social problems. An additional seven security items (A to G) aim to track current clinical risk management needs related to harm to others, harm to self, vulnerability, the built environment, and staffing (see Table 1). The authors state that neither set of items is intended to constitute a risk assessment tool and should not be substituted for one (Sugarman & Walker, 2007); rather, both were conceptualised to track a range of health- and social-related clinical and risk outcomes in terms of need for secure care based on recent clinical assessments including, but not limited to, risk assessments. The HoNOS-Secure was introduced into the current study setting in 2004 as a routine outcome measure. Both sets of items have acceptable inter-rater reliability in routine clinical practice (Dickens et al., 2007) and can detect statistically significant change over time when used to calculate a performance metric in patient cohorts (Sugarman et al., 2009). The policy guideline is that: i) each patient is to be rated at three-monthly intervals, commencing at the pre-admission assessment; ii) that rating should be undertaken by the multidisciplinary team responsible for the patient's care; and iii) that raters have received training. These guidelines are overseen at board level through key performance indicators. Each HoNOS-secure item is rated on a five-point scale (0 to 4) with each point representing an anchor with an accompanying narrative description. Because rating is a regular part of clinical practice there is no blinding to outcome, and raters are aware that data will be used for evaluation purposes.

>>Insert Table 1 about here<<

Patient characteristics. Patients' age at assessment, gender, admission and discharge dates, legal status, security level, self-reported ethnicity, and ICD-10 (World Health Organisation, 1992) psychiatric diagnoses were extracted from patient records. A large proportion of patients had both schizophrenia and personality disorder; therefore, they were treated as a distinct group in the analysis.

Data Analysis

Descriptive statistics were conducted to examine the distribution of sample characteristics and HoNOS-Secure item and total scores. Cronbach's alpha was calculated for items 1 to 12 and the security items A to G as measures of internal reliability; scores in the range of .7 to .9 are viewed as acceptable (Tavakol & Dennick, 2011). In order to examine differences in baseline scores and rates of change in HoNOS-Secure item and total scores across demographic and clinically defined groups, linear mixed model analyses with maximum likelihood estimation (MLE; θ) were used; whether time between assessments affected rates of change was also examined using this method. Mixed effects models are preferred as they allow for modelling repeated assessments with a random effect to account for variation among patients' scores (Antonio & Beirlant, 2007) and can account for differing numbers of assessments per patient. Fixed effects were used to model variables of interest with the resulting estimates representing the mean difference between the scores of those in the group of interest relative to those in the reference category after controlling for covariates; the relevant reference categories were: gender (reference category = male), age (reference category = aged less than 39 years), ethnicity (reference category = Caucasian), diagnosis (reference category = schizophrenia), substance abuse (reference category = no), Mental Health Act status (reference category = civil), and security level of ward (reference category = low secure). Fixed effects were also estimated for the interaction between assessment iteration and each of the factors to examine rates of change relative to the reference group.

The resulting *t*-values and accompanying degrees of freedom are calculated by default in SPSS using the Satterthwaite approximation (Satterthwaite, 1946) which does not assume equal variances between samples and was developed for use in cases where only estimates of the variance are known; in such cases it is possible for degrees of freedom to be fractional. In order to counteract the increased probability of Type I error from testing of all 19 HoNOS-secure items plus the two total scores on each variable of interest a Bonferroni correction was applied (α=.05/ *n*=21 hypotheses = .002). Cohen's *d* values were calculated to examine the within subjects effect size of changes in scores and between groups differences in baseline scores; 0.2, 0.5, and 0.8 are typically considered the thresholds for small, medium, and large effect sizes (Cohen, 1992).

Clinically significant change is change that has taken the person from a score typical of a problematic, dysfunctional, patient, client, or user group to a score typical of the "normal" population (Jacobson & Truax, 1991). Scores of 2 or above on any of the HoNOS items1-12 indicates that a treatment intervention is required; scores of 1 or above on the security items A-G indicate requirement for risk-management interventions (Long et al., 2010); therefore, a change in scale score from at or above the corresponding cut-off value to below was identified as a clinically significant positive change; and a change in score from below the cut off to the cut off or above was identified as clinically significant negative change, i.e., deterioration. These were calculated as proportions of the total cohort size along with those of patients who had been at or above the cut off at first assessment and remained as such at the last assessment (i.e., remained problematic); and those who were below the cut off at both points (i.e., remained non-problematic).

The Reliable Change Index (RCI; Evans, Margison, & Barkham, 1998), which examines whether change is greater than would be expected to happen based on measurement error alone, was calculated to determine whether the difference between the follow-up and

initial total scale scores exceeds a threshold calculated as a function of the initial standard

deviation of the measure and its reliability (see formula below; SD_1 = standard deviation of

baseline scores on the measure, r = reliability of the measure).

$$SE_{diff} = SD_1\sqrt{2\sqrt{1} - r}$$

The use of Cronbach's alpha or another parameter of internal consistency is the most

theoretically consistent approach since a test-retest reliability measure always includes not

only simple unreliability of the measure but also any real changes in whatever is being

measured. This means that internal reliability is almost always higher than test-retest and will

generally result in more people being seen to have changed reliably. Therefore, the

Cronbach's alphas derived from HoNOS ratings in the current study, were used as the

measures of reliability. Finally, we calculated the proportion of patients who had undergone

reliable change and clinically significant change on an item-by-item basis. All statistics were

conducted using PASW Statistics version 18 for Windows.

Results

Sample Characteristics

There were 467 individuals who met the inclusion criteria. Individuals were classified

as i) having a shorter interval between assessments if any interval between pairs of

assessments was less than the mean interval between assessments minus two standard

deviations, ii) having a longer interval between assessments if any interval between

assessments was greater than the mean plus two standard deviations, or iii) not deviating

from the mean if the interval between all assessments was within two standard deviations of

the mean. Entering this as a factor revealed that rates of change of five items was affected,

namely: Item 6, $F(10, 3818) = 1.87$, $p=.04$; item 11, $F(10, 1940) = 2.05$, $p=.03$; item 12,

$F(10, 1930) = 2.49$, $p=.006$; item C, $F(10, 3283) = 2.36$, $p=.009$; item E, $F(10, 1911) = 1.99$,

p=.03). Therefore, individuals whose interval between assessments deviated from the mean were excluded from further analysis, leaving a final sample size of 418.

The HoNOS-Secure assessments spanned a mean 726.6 days (SD=151.9) per patient; the mean interval between admission and first assessment was 799.0 days (SD=1381.9). Approximately two thirds of the sample were male (n=281; 67.2%); mean age at first assessment was 38.7 years (SD=14.5), which was classified as aged 39 years or above (n=183; 43.8%) and aged less than 39 years (n=235, 56.2%) for further analysis. There were 192 Caucasian patients (45.9%), 37 Black patients (8.9%), 19 Asian patients (4.5%), 19 patients with a mixed ethnic background (4.5%), 2 patients from an "other" ethnic background (0.5%), and 149 patients (35.6%) with no recorded ethnicity; this was grouped as Caucasian (n=192; 45.9%), non-Caucasian (n=77; 18.4%), and unknown (n=149; 35.6%) for analysis. Three quarters of the sample (n=321; 76.8%) were resident in low secure wards and one quarter (n=97; 23.2%) were in conditions of medium security.

The most common ICD-10 (World Health Organisation, 1992) psychiatric diagnoses were Schizophrenia, schizotypal and delusional disorders (F20-29; n=256, 61.2%), Disorders of adult personality and behaviour (F60-69; n=156, 32.5%), Disorders of psychological development (F80-89; n=60, 14.4%), Mental and behavioral disorders due to psychoactive substance use (F10-F19; n=61; 14.6%), Mood [affective] disorders (F30-39; n=28; 6.7%) and Organic, including symptomatic, mental disorders (F00-09; n=17, 4.1%). Substance abuse was entered as a separate variable, coded as present or absent, to minimise the level of complex co-morbidity in the sample. For analysis, participants were categorised according to the remaining ICD-10 major categories based on diagnosis within a single ICD-10 category. There were, however, a large number of patients with both schizophrenia and personality disorder; therefore, these were treated as a distinct group. There were 74 patients (17.7%) with multiple or uncommon diagnoses that could not be grouped for analysis in any

meaningful way. Therefore, the final diagnostic groups were schizophrenia (*n*=198, 47.4%),

personality disorder (*n*=67, 16.0%), schizophrenia & personality disorder (*n*=36, 8.6%),

developmental disorder (*n*=31, 7.4%), organic disorders (*n*=12, 2.9%), and other or multiple

diagnoses (*n*=74, 17.7%).

HoNOS-Secure Assessments

The mean interval between assessments was 141.0 days (SD=62.8). There were 25

individuals (6.0%) who only had two assessments, 55 had three (13.2%), 37 (8.9%) had four,

47 (11.2%) had five, and 254 (60.8%) had six assessments completed. The mean scales and

total scores for each assessment are presented in Table 2. Cronbach's alpha for scales 1-12

was .722, and for scales A to G was .714. The HoNOS scales 1 to 12 total score decreased

from assessment one to three, but this was not maintained; total security scales score

decreased from assessment one to five but increased at the final assessment.

>>Insert Table 2 about here<<

Baseline Scores

HoNOS-Secure 1 to 12 total scores. Analysis revealed some significant between

group baseline total score differences (see Table 3). Baseline scores were significantly higher

in medium secure patients compared with those resident in low secure wards. Women had

significantly higher total scores relative to men. Baseline total scores were lower for

individuals admitted informally compared to individuals admitted under civil sections of the

Mental Health Act, as were scores on items 1, 4, 6, 9, and 10 (range 0.47 to 0.88).

HoNOS-Secure 1 to 12 item scores. Patients in medium secure wards had

significantly higher scores than those in low secure wards on item 1, θ=.56, t(457.20) = 5.85,

p<.001, 95% CI [.37, .74] and item 9, θ=.43, t(526.58) = 4.21, *p*<.001, 95% CI [.23,.62].

Women had significantly higher scores than men on item 2, θ=.19, t(681.13) = 2.50, *p*<.001,

95% CI [.04,.34]; item 5, θ=.42, t(630.48) = 4.55, *p*<.001, 95% CI [.22,.62]; item 7, θ=.18,

t(462.16) = 2.43, *p*<.001, 95% CI [.32, .63]; and item 8, ɵ=.58, *t*(538.11) = 5.77, *p*<.001, 95%

CI [.38, .78].

Those with an organic diagnosis had higher scores than individuals with

schizophrenia on item 4 (ɵ=1.01, *t*(341.25) = 4.16, p<.001, 95% CI [.5, 1.48], item 5 (ɵ=1.02,

t(623.89) = 3.99, p<.001, 95% CI [.52, 1.52], and item 11, ɵ=.48, *t*(520.72) = 3.37, p=.001,

95% CI [.20, .76], and a significantly lower score on item 6, ɵ=-1.04, *t*(467.67) = -3.96,

p<.001, 95% CI [-1.69, -.57]. Individuals aged 39 years or above had a higher score on item

5, ɵ=.45, *t*(627.99) = 4.10, p<.001, 95% CI [.28, .62] than younger patients. Those with a

substance abuse diagnosis scored higher compared with those without a substance abuse

diagnosis on the related item 3, ɵ=.15, *t*(860.38) = 4.38, p<.001, 95% CI [.08, .22]. Scores for

individuals with schizophrenia for item 6 were higher than for those with a personality

disorder, ɵ=1.60, *t*(471.78) = 10.10, p<.001, 95% CI [1.29, 1.91]; those with co-morbid

schizophrenia and personality disorder, ɵ=1.39, *t*(341.25) = 4.16, p<.001, 95% CI [1.01,

1.76]; those with a developmental diagnosis ɵ=1.13, *t*(467.67) = 3.96, p<.001, 95% CI [.57,

1.69]; those with an organic disorder ɵ=.74, *t*[470.20]=4.26; *p*<.001), 95% CI [.40, 1.08]; and

those with multiple or "other" disorders, ɵ=1.04, *t*[471.24]=7.89; *p*<.001, 95% CI [.78. 1.30].

Individuals with personality disorder also scored higher than individuals with schizophrenia

on item 2, ɵ=0.76, *t*[693.04]=7.04; *p*<.001, 95% CI [.55, .97], and item 7, ɵ= 0.35,

t[469.94]=3.31; *p*=.001, 95% CI [.14, .56], but they had lower scores on items 4, ɵ = 0.77,

t[353.38]=-5.69; *p*<.001, 95% CI [.50, 1.03], and item 10, ɵ = 0.52 , *t*[471.82]=3.36; *p*=.001,

95% CI [.22, .82]. Those with developmental disorders had higher scores relative to those

with schizophrenia on item 9, ɵ= 0.71, *t*[529.49]=4.13; *p*<.001, 95% CI [.37, 1.04]. Finally,

individuals with multiple or "other" diagnoses scored higher on item 2, ɵ =.46,

t[692.16]=5.10; *p*<.001, 95% CI [.28, .63], and item 7, ɵ =.40, *t*[469.19]=4.38; *p*<.001, 95%

CI [.22, .57] relative to those with schizophrenia.

HoNOS-Secure A to G total scores. Patients resident in medium secure wards had

higher baseline security items total score relative to low secure patients ɵ =4.28,

t[478.45]=12.85; *p*<.001, 95% CI [3.62, 4.93]. Women had higher baseline total score when

compared with men, ɵ =1.26, *t*[481.42]=3.73; *p*<.001, 95% CI [.60, 1.93]. Patients admitted

informally had lower total security score compared with patients admitted under civil sections

of the Mental Health Act, ɵ =-2.14, *t*[514.37]=-4.33; *p*<.001 95% CI [-1.16, -3.12].

HoNOS-Secure A to G item scores.

Compared with individuals resident on low secure wards, those individuals resident on

medium secure wards had significantly higher scores on item A, ɵ =.65, *t*[522.19]=6.23;

p<.001, 95% CI [.44, .85]; item C, ɵ =.92, *t*[420.98]=17.55; *p*<.001, 95% CI [.82, 1.03]; item

D, ɵ =.51, *t*[479.67]=8.63; *p*<.001, 95% CI [.39, .62]; item E, ɵ =1.01, *t*[420.92]=11.88;

p<.001 95% CI [.85, 1.18]; and item G, ɵ =.72, *t*[496.92]=9.64; *p*<.001, 95% CI [.58, .87].

Women had significantly higher scores than men on item B, ɵ =.58, *t*[523.73]=2.73; *p*<.001;

95% CI [.38, .79], and on item F, ɵ =.55, *t*[530.26]=6.55; *p*<.001, 95% CI [.38, .71].

Individuals with a sole diagnosis of schizophrenia scored lower on item B than those with a

personality disorder, ɵ =.95, *t*[530.35]=6.46; *p*<.001, 95% CI [.66, 1.24]. Those admitted

under forensic sections of the Mental Health Act had higher baseline scores relative to those

admitted under civil sections on item A, ɵ =.64, *t*[524.97]=6.83; *p*<.001, 95% CI [.46, .82];

however, they had a lower score on item F, ɵ =-.25, *t*[534.02]=3.41; *p*<.001, 95% CI [-.11, -

.40]. Individuals with a developmental diagnosis scored lower on item G than those with

schizophrenia, ɵ =-.41, *t*[491.14]=-3.19; *p*<.001, 95% CI [-.16, -.65]. Individuals aged 39

years and above had a lower score on item E, ɵ =-.24, *t*[425.66]=-.514; *p*=.001, 95% CI [-.10,

-.38] relative to those aged less than 39 years. Finally, patients admitted informally scored

lower than the reference category (civil section) on item C, θ =-.29, t[423.19]=-2.10; p<.001

95% CI [-.14, -.45]; item D, θ =-.32, t[538.28]=-3.57; p<.001, 95% CI [-.14, .49]; and item E,

θ =-.54, t[464.00]=-4.25; p<.001, 95% CI [-.29, -.79].

>>Insert Tables 3 and 4 about here<<

Rates of Change

      HoNOS items 1 to 12. The mean total score reduced by 0.37 points per assessment,

though not significantly so after Bonferroni correction, t[118.96]=-2.20; p=.030; none of the

individual items reduced significantly across repeated assessments after correction (see Table

3). Individuals with co-morbid schizophrenia and personality disorder increased relative to

individuals with a sole diagnosis of schizophrenia on item 8 score, θ = 0.18, t[142.26]=3.26;

p=.001, 95% CI [.07, .29].

>>Insert Table 5 about here<<

      Security items. The security items total score reduced significantly across repeated

assessments, θ =-0.40, t[166.63]=-3.62; p<.001, 95% CI [-.18, -.62], as did items D and E (see

Table 5).

Clinically Significant and Reliable Change

      Table 4 shows the proportions of patients undergoing clinically significant change

over the assessment period, that is moving from the problematic to the non-problematic range

or vice versa. On the individual scales 1 to 12, clinically significant improvement was

revealed for between 1.9% (item 3) and 16.5% (item 8) of patients and corresponding

deterioration for between 0.5% (item 3) and 16.0% (item 8) of patients; corresponding figures

for the individual security scales were improvements for between 0.2% (item G) and 12.9%

(item E), and deterioration for between 0.2% (items C,D,G) and 6.5% (item B). . The mean

number of HoNOS items on which individuals' scores were in the normal range before and

after the assessments was 7.28 (*SD*=2.18); the mean numbers of items on which scores

improved, worsened, and stayed in the problematic range were 1.38 (*SD*=1.40), 1.00 (*SD*=1.19), and 2.35 (*S*=1.79), respectively.

There were 20 (4.8%) patients that showed a statistically reliable reduction in total HoNOS scale 1 to 12 scores; however, 6 (1.4%) individuals actually showed a reliable increase on HoNOS scores, indicating global deterioration. Of the 20 individuals that underwent reliable reductions, clinically significant change on each item occurred for between 2 (0.5%) and 11 (2.6%) of patients; the greatest number of individuals reduced on item 2 and item 7, whereas item 3 showed the least reduction

>>Insert Table 6 about here<<

For the HoNOS-Secure security scales A to G, 24 (5.7%) individuals showed reliable reductions in total score, but 2 (0.2%) individuals had a reliable increase in scores. Of the 24 that reliably reduced, no-one showed clinically significant change on item D or item G, and only one person demonstrated clinically significant change on item C; however, five to twelve individuals reduced significantly on the remaining items, the greatest number being for item E. The mean number of security items on which individuals scores were in the normal range before and after the assessments was 0.26 (*SD*=0.53); the mean numbers of items on which scores improved, worsened, and stayed in the problematic range were 0.44 (*SD*=0.81), 0.17 (*SD*=0.50), and 6.13 (*S*=1.19), respectively.

Discussion

In the current study HoNOS-Secure assessments captured statistically significant baseline differences between clinically and diagnostically defined groups; differences were mostly in accord with what we might predict from the existing evidence. While it is hardly surprising that ratings differed between certain clinical groups in predictable ways – for example, higher scores suggestive of more global and risk related impairment were found among those detained in conditions of medium security compared with those in low security

– the finding does suggest a crude indicator of face validity of the HoNOS-Secure. This also

occurred despite the fact that rating was conducted in routine practice by different clinical

teams across time which could also support the view that the HoNOS-Secure scales have

sufficient ecological validity to detect predictable between-group differences in the real

world.

However, the purpose of the HoNOS-Secure is not as an assessment tool and these

broadly positive findings are of far less importance than findings about i) whether *rate* of

change, rather than baseline assessment score, differs between clinical and demographic

groups; and ii) the ability of the scales to detect the changes in risk-related need, behaviour,

symptomatology, and functioning that we might reasonably expect as a result of exposure to

highly resource-intensive services. Regarding the first aim: statistically significant differences

in rates of change between clinical and demographic groups was limited to a single finding

about worsening scores on HoNOS-secure item 8 (Other mental and behavioural problems)

for patients with diagnoses of schizophrenia and personality disorder relative to those with

schizophrenia and no personality disorder. This is consistent with suggestions that people

with personality disorder in forensic care may be susceptible to iatrogenic treatment effects

(e.g., Linehan, 1993), although it does not explain why those solely with a personality

disorder diagnosis did not react similarly. In the case of the second aim, analyses suggested

that HoNOS 1 to 12 total score fell, indicating improvement (small effect size), over time,

though not to a statistically significant degree taking multiple testing into account. The

security scales total score fell significantly (medium effect size), again indicating

improvement, across the period of assessment; items about the need for a staffed living

environment and need for an escort on leave fell significantly over time. Despite this, mean

item differences were all so small as to be, in all likelihood, clinically negligible. This means

that, in effect, overall findings about the rate of change, or lack thereof, in the inpatient

setting are generalizable across these clinical and demographic groups. Our finding that a very small number of people changed reliably in a negative direction, and that just 4.3% and 5.1% changed reliably on total scores on the HoNOS scales 1 to 12 and security scale totals respectively in a positive direction is disappointing and raises a number of important questions.

First, we should ask whether the lack of change simply reflects that many patients did not fall into the problematic category in the first instance and thus had little room for improvement over time. This may play a role in the explanation of score stasis on the HoNOS scales 1 to 12 where between 17.5% and 97.4% (median=61.7%) patients were in the non-problematic range at first and final assessment. The current study design did not allow us to explore whether these patients had undergone improvements during admission prior to the first assessments presented here, or whether their non-problematic status preceded their admission. This is an acknowledged limitation of the current study. As routine collection of outcomes data continues it may become possible to further examine this question. However, there was clinically significant change across the assessment period; between 0.5% and 16.0% (median 10.6%) deteriorated on items 1 to 12, and between 1.9% and 16.5% (median 12.9%) improved. Further, between 0.2% and 60.3% (median 14.1%) remained unchanged in the problematic category; again, it could be that any improvement these patients had made was prior to commencement of assessment data included in this study, however, the further period of admission had not improved outcomes to a clinically significant degree. For the security scales, many patients' risk-related outcome remained static, though in this case stasis was accounted for by problematic ratings for items at both assessment points (70.3% to 99.6%, median = 86.4%). More patients improved (0.2% to 12.1%, median 6.9%) than deteriorated (0.2% to 6.5%, median 1.2%).

While it is disappointing that there was relatively little clinically significant change, very little statistically reliable change, and even some statistically reliable deterioration, the current study design does not allow us to definitively conclude that this is indicative of inadequate or ineffective care and treatment from such relatively well-resourced services, or whether it is suggestive of an inadequate outcomes measurement tool. Relevant, related research is scant but does suggest that reliable and clinically significant change in populations like that in the current study, and for outcomes similar to those studied here, is not captured by existing routinely used tools. For example, we (O'Shea & Dickens, 2015) have previously reported that just 3% of individuals in the same secure mental health service as in the current study had clinically significant and reliable reduction in risk, as measured on routine applications of the HCR-20 Version 2.0 (Webster, Douglas, Eaves, & Hart, 1997). The similarly small proportion of patients making significant progress in perceived risk and risk-related need leads us to question whether the broad and complex range of interventions and management strategies employed in secure services are warranted by their routinely demonstrated outcomes. Of course, policymakers may be interested solely in long term outcomes following discharge including recidivism, and mortality; there is limited research from long term follow-up studies. In the UK, of 550 patients discharged from a medium secure care facility, adverse outcomes over a 20-year period included death (10%; 18% of these by suicide), reconviction (49%), and readmission to psychiatric care (38%) (Davies, Clarke, Hollin & Duggan, 20007).

Perhaps of most concern is that mean scores for those remaining in services over repeated HoNOS-Secure iterations began to rise. This may indicate overall deterioration in levels of behaviour, symptomatology, and functioning at iteration six following promising reductions at earlier assessments. The results suggest that that the growth in the number, diversity, and complexity of secure beds in the UK, particularly at levels of medium and low

security (Sugarman & Dickens, 2015), is not paralleled by evidence of the benefits of

significant positive clinical change.

Limitations

The current study was conducted in only one provider of secure services and it is therefore

possible that the changes, or rather lack of change, detected may not be mirrored in other

provider units. Further, the HoNOS-Secure has been subject to limited research regarding its

validity. This does mean that our speculation about the value of therapeutic input in secure

services remains merely speculation, though we would be surprised if substantially different

results would be found in other units or using other routinely administered measures.

Astute readers will have noticed the apparently long mean (799 days) period of time between

patient admission and first HoNOS-Secure assessment. This should not be taken as evidence

of an unnecessary delay in making an assessment; rather, it is reflective of the fact that many

individuals were existing patients before the introduction of the HoNOS-Secure as a

mandated tool in the study setting. In fact, risk assessment for these patients occurred as soon

as practically possible following the introduction of the tool but we have chosen to present

correct information about time from admission to assessment rather than from introduction of

the tool to assessment.

Conclusions

In the main, very few patients in this typical secure mental health service underwent reliable

or clinically significantly improvement in terms of risk-related need, functioning, behaviour

or symptomatology over time that was detected by the HoNOS-Secure. This mirrors previous

research about lack of change in perceived risk (O'Shea & Dickens, 2015). The taxpaying

public and users of services should rightfully question what the purpose of secure care is if

not to facilitate, and subsequently demonstrate, positive change on these important factors.

REFERENCES

Antonio, K., & Beirlant, J. (2007). Actuarial statistics with generalized linear mixed models. *Insurance: Mathematics and Economics, 40*, 58-76. doi: 10.1016/j.insmatheco.2006.02.013

Audin K., Margison F., Clark, J.M. & Barkham, M. (2001). Value of HoNOS in assessing patient change in NHS psychotherapy and psychological treatment services. *British Journal of Psychiatry, 178*, 561-566. doi: 10.1192/bjp.178.6.561

Baxter, R., Rabe-Hesketh, S., & Parrott, J. (1999). Characteristics, needs and reoffending in a group of patients with schizophrenia formerly treated in medium security. *Journal of Forensic Psychiatry and Psychology, 10*, 69-83. doi:10.1080/09585189908402140

Buchanan, A. (1998). Criminal conviction after discharge from special (high security) hospital. Incidence in the first 10 years. *British Journal of Psychiatry, 172*, 472-476. doi: 10.1192/bjp.172.6.472

Burckhardt, C. S., & Anderson, K. L. (2003). The Quality of Life Scale (QOLS): Reliability, validity, and utilization. *Health and Quality of Life Outcomes, 1*, 60. http://doi.org/10.1186/1477-7525-1-60

Burgess, P., Pirkis, J. & Coombs, T. (2006). Do adults in contact with Australia's public health mental health services get better? *Australia and New Zealand Health Policy, 3*, 9. doi: 10.1186/1743-8462-3-9.

Chambers, J, Yiend, J., Barrett, B., Burns, T., Doll, H., Fazel, S.,… Fitzpatrick, R. (2009). Outcome measures used in forensic mental health research: a structured review. *Criminal Behaviour and Mental Health, 19*, 9-27. doi: 10.1002/cbm.724

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112*, 155. doi: http://dx.doi.org/10.1037/0033-2909.112.1.155

Coid, J Hickey, N, Kahtan, N. Zhang, T & Yang, M. (2007). Patients discharged from medium secure forensic psychiatry service: reconvictions and risk factors. *British Journal of Psychiatry, 190*, 223-229. doi: 10.1192/bjp.bp.105.018788

Davies, S., Clarke, M., Hollin, C & Duggan, C. (2007). Long-term outcomes after discharge from medium secure care: a cause for concern. *British Journal of Psychiatry, 191*, 70-74. doi: 10.1192/bjp.bp.106.029215

Delaffon, V., Anwar, Z., Noushad, F., Ahmed, A.S., & Brugha, T.S. (2012). Use of Health of the Nation Outcome Scales in psychiatry. *Advances in Psychiatric Treatment, 18*, 173–179. doi: 10.1192/apt.bp.110.008029

Department of Health (2007). *Best Practice Guidance: Specification for adult medium-secure services*. Retrieved from: http://webarchive.nationalarchives.gov.uk/20130107105354/http://www.dh.gov.uk/prod_consum_dh/groups/dh_digitalassets/documents/digitalasset/dh_078050.pdf

Dickens, G., Sugarman, P., & Walker, L (2007). HoNOS-Secure: a reliable outcome measure for users of secure and forensic mental health services. *Journal of Forensic Psychiatry and Psychology, 18*, 507-514. doi: 10.1080/14789940701492279

Dickens, G., Picchioni, M., Long, C., & Sugarman, P. (2010). HoNOS-Secure: assessing risk and recovery for men in secure care. *British Journal of Forensic Practice, 12*, 36-45. doi: http://dx.doi.org/10.5042/bjfp.2010.0613

Edwards, J., Steed, P., & Murray, K. (2002). Clinical and forensic outcome 2 years and 5 years after admissions to a medium secure unit. *Journal of Forensic Psychiatry, 13*, 68-87. doi: 10.1080/09585180210123294

Ellwood, P. M. (1988). Outcomes management. *New England Journal of Medicine, 318*, 1549-1556. doi: 10.1056/NEJM198806093182329

Evans, C., Margison, F., & Barkham, M. (1998). The contribution of reliable and clinically significant change methods to evidence-based mental health. *Evidence-Based Mental Health, 1*, 70-72. doi:10.1136/ebmh.1.3.70

Fitzpatrick, R., Chambers, J., Burns, T., Doll, H., Fazel, S., Jenkinson, C., Kaur, A., Knapp, M., Sutton, L., & Yiend, J. (2010). A systematic review of outcome measures used in forensic mental health research with consensus panel opinion. *Health Technology Assessment,14,* 18. doi: 10.3310/hta14180

Hall, R.C. (1995) Global Assessment of Functioning. A modified scale. *Psychosomatics, 36*, 267-275. doi: doi: http://dx.doi.org/10.1016/S0033-3182(95)71666-8

Holloway F (2002). Mentally disordered offenders and the community mental health team. In A. Buchanan(Ed.), *Care of the mentally disordered offender in the community* (pp. 222–242). Oxford: Oxford Medical Publications.

Hunter, R., Cameron, R. & Norrie, J. (2009). Using patient reported outcomes in schizophrenia: The Scottish schizophrenia outcomes study. *Psychiatric Services, 60*, 240-245. doi: 10.1176/appi.ps.60.2.240

Jacobson, N.S. & Truax, P. (1991). Clinical significance: a statistical approach to defining meaningful change in psychotherapy research. *Journal of Consulting and Clinical Psychology, 59*,12-19. doi:10.1037/0022-006x.59.1.12

Linehan, M. (1993) *Cognitive behavioural treatment of boderline personality disorder*. New York: Guldford.

Long, C. G., Dickens, G., Sugarman, P., Craig, L., Mochty, U., & Hollin, C. (2010). Tracking Risk Profiles and Outcome in a Medium Secure Service for Women: Use of the HoNOS-Secure. *International Journal of Forensic Mental Health, 9*, 215-225. doi: 10.1080/14999013.2010.517256

Maden, A., Scott, F., Burnett, R., Lewis, G.H., & Skapinakis, P. (2004). Offending in psychiatric patients after discharge from medium secure units: prospective national cohort study. *BMJ, 328*, 1534. doi: http://dx.doi.org/10.1136/bmj.38075.467569.EE

Maden, A., Skapinakis, P., Lewis, G., Scott, F., Burnett, R. & Jamieson, E. (2006). Gender differences in reoffending after discharge from medium secure units: national cohort study in England and Wales. *British Journal of Psychiatry, 189*, 168-171. doi: 10.1192/bjp.bp.105.014613

Markovitz, P. (1996). *The Avon Mental Health Measure*. Bristol: Changing Minds.

O'Shea, L.E. & Dickens, G.L. (2015). The HCR-20 as a measure of reliable and clinically significant change in violence risk among secure psychiatric inpatients. *Comprehensive Psychiatry, 62*, 132-140. doi: http://dx.doi.org/10.1016/j.comppsych.2015.07.009

Pirkis, J., Burgess, P., Coombs, T. Clarke, A., Jones-Ellis, D., & Dickson, R. (2005). Routine measurement of outcomes in Australia's public sector mental health services. *Australia and New Zealand Health Policy, 2*, 8. doi: 10.1186/1743-8462-2-8.

Rutherford, M. & Duggan, S. (2007). *Forensic mental health services. Facts and figures on current provision*. London: Sainsbury Centre for Mental Health.

Sahota, S., Davies, S., Duggan, C., and Clarke, M. (2009). The fate of medium secure patients discharged to generic or specialised services. *Journal of Forensic Psychiatry and Psychology, 20*, 74-84. doi:10.1080/14789940802221304

Satterthwaite, F.E. (1946) An approximate distribution of estimates of variance components. *Biometrics Bulletin, 2*, 110-114. doi: 10.2307/3002019

Sugarman, P. & Dickens, G.L. (2015). The evolution of secure and forensic mental health care. In G. Dickens, P. Sugarman, & M. Picchioni (Eds.), *Handbook of Secure Care* (pp.1-14). London: RCPsych Publications.

Sugarman, P. & Walker, L.(2007). *HoNOS-Secure version2b* http://www.rcpsych.ac.uk./research& training unit/HoNOS/secure.aspx

Sugarman, P., Walker, L. & Dickens, G. (2009). Managing outcomes performance in mental health using HoNOS: Experience at St Andrew's Healthcare. *Psychiatric Bulletin, 33*, 285-288. doi: 10.1192/pb.bp.108.022475

Sugarman, P. & Dickens, G. (2010). Demonstrating positive global and risk-related outcomes in mental health with HoNOS. *British Journal of Healthcare Management, 16*, 431-435. doi: 10.12968/bjhc.2010.16.9.78369

Tavakol, M., Dennick, R. (2011) Making sense of Cronbach's alpha. *International Journal of Medical Education, 2*, 53-55. doi: 10.5116/ijme.4dfb.8dfd

Turner, M.A., Boden, J.M., Smith-Hamel, C., Mulder, R.T. (2009). Outcomes for 236 patients from a 2-year early intervention in psychosis service. *Acta Psychiatrica Scandinavica, 120*, 129-137. doi: 10.1111/j.1600-0447.2009.01386.x

Webster, C. D., Douglas, K. S., Eaves, D., & Hart, S. D. (1997). *HCR-20: Assessing the Risk for Violence (Version 2)*. Vancouver: Mental Health, Law, and Policy Institute,Simon Fraser University.

Wing, J.K., Beevor, A.S., Curtis, R.H., Park, S.B., Hadden, S., & Burns, A. (1998). Health of the Nation Outcome Scales: Research & Development. *British Journal of Psychiatry, 172*, 11–18. doi: 10.1192/bjp.172.1.11

World Health Organisation. (1992). *The ICD-10 classification of mental and behavioural disorders*. Geneva: WHO.

Table 1: HoNOS-Secure (Version 2b) scale items

Scales 1 to 12:

1. Overactive, aggressive, disruptive or agitated behaviour
2. Non-accidental self-injury
3. Problem drinking or drug taking
4. Cognitive problems
5. Physical illness or disability problems
6. Problems with hallucinations and delusions
7. Problems with depressed mood
8. Other mental and behavioural problems
9. Problems with relationships
10. Problems with activities of daily living
11. Problems with living conditions
12. Problems with occupations and activities

Security Scales:

A. Risk of harm to adults or children
B. Risk of self-harm
C. Need for building security to prevent escape
D. Need for a safely-staffed living environment
E. Need for escort on leave
F. Risk to individual from others
G. Need for risk management procedures

Table 2: Distribution of HoNOS-Secure scale and total scores as a function of assessment

Assessment number	Mean (SD) score					
	1 n=418	2 n=418	3 n=393	4 n=338	5 n=301	6 n=254
Scales 1 to 12						
1	1.19 (0.91)	1.12 (0.96)	1.11 (1.00)	1.10 (0.97)	1.08 (0.95)	1.07 (0.94)
2	0.45 (0.90)	0.38 (0.82)	0.36 (0.81)	0.31 (0.76)	0.33 (0.78)	0.27 (0.73)
3	0.07 (0.40)	0.03 (0.28)	0.01 (0.12)	0.02 (0.19)	0.03 (0.21)	0.02 (0.27)
4	0.88 (0.97)	0.85 (0.96)	0.82 (0.92)	0.85 (0.93)	0.83 (0.91)	0.86 (0.92)
5	1.00 (1.01)	1.01 (1.02)	0.93 (0.95)	0.97 (0.95)	0.96 (0.94)	0.95 (0.95)
6	1.26 (1.24)	1.22 (1.24)	1.19 (1.26)	1.16 (1.24)	1.13 (1.21)	1.15 (1.19)
7	0.98 (0.84)	0.92 (0.75)	0.90 (0.81)	0.87 (0.77)	0.88 (0.79)	0.89 (0.81)
8	1.76 (1.04)	1.71 (0.99)	1.63 (0.99)	1.71 (1.04)	1.71 (1.00)	1.77 (0.94)
9	2.05 (1.05)	2.01 (1.02)	1.97 (1.01)	2.01 (0.99)	1.90 (0.98)	2.00 (1.02)
10	1.50 (1.08)	1.45 (1.09)	1.39 (1.08)	1.38 (1.08)	1.40 (1.08)	1.49 (1.07)
11	0.51 (0.62)	0.46 (0.53)	0.46 (0.54)	0.46 (0.55)	0.45 (0.55)	0.39 (0.54)
12	0.52 (0.70)	0.41 (0.57)	0.35 (0.55)	0.33 (0.54)	0.33 (0.52)	0.31 (0.53)
Total	12.16 (5.43)	11.57 (5.11)	11.13 (5.00)	11.17 (5.05)	11.02 (4.94)	11.17 (5.16)
Security scales						
A	2.10 (1.09)	2.06 (1.05)	1.98 (1.07)	1.99 (1.09)	2.01 (1.08)	1.96 (1.11)
B	1.45 (1.12)	1.39 (1.06)	1.35 (1.08)	1.25 (1.06)	1.29 (1.10)	1.34 (1.12)
C	2.13 (0.59)	2.12 (0.63)	2.09 (0.68)	2.10 (0.68)	2.06 (0.72)	2.10 (0.69)
D	2.10 (0.58)	2.03 (0.58)	2.01 (0.63)	1.96 (0.60)	1.94 (0.62)	1.94 (0.63)
E	1.84 (0.90)	1.62 (0.89)	1.58 (0.89)	1.55 (0.90)	1.46 (0.87)	1.46 (0.93)
F	1.52 (0.83)	1.54 (0.80)	1.50 (0.78)	1.45 (0.82)	1.48 (0.80)	1.48 (0.79)
G	2.21 (0.79)	2.23 (0.80)	2.27 (0.76)	2.30 (0.75)	2.27 (0.76)	2.34 (0.72)
Total	13.34 (3.66)	13.00 (3.68)	12.78 (3.89)	12.60 (3.89)	12.50 (3.94)	12.63 (3.91)

Table 3: Estimates of fixed effects for HoNOS-Secure items 1 to 12 total score

Parameter	Estimate	Std. Error	df	t	Significance	95% CI
Intercept	12.47	.54	532.38	23.23	<.001	11.42-13.52
Security Level = Medium	2.06	.52	525.53	3.96	<.001	1.04-3.08
Security Level = Low	0	0	-	-	-	-
Gender=Female	2.12	.53	530.82	4.01	<.001	1.08-3.16
Gender=Male	0	0	-	-	-	-
Legal Status=Informal	-4.08	.78	582.65	-5.24	<.001	-5.60 - -2.55
Legal Status=Forensic	-3.54	.47	528.80	-7.56	<.001	-4.46 - -2.62
Legal Status= Civil	0	0	-	-	-	-

Table 4: Estimates of fixed effects for HoNOS-Secure items A to G total score

Parameter	Estimate	Std. Error	df	t	Significance	95% CI
Intercept	11.92	.34	482.43	34.68	<.001	11.24-12.60
Security Level = Medium	4.28	.33	478.45	12.85	<.001	3.62-4.93
Security Level = Low	0	0	-	-	-	-
Gender=Female	1.26	.34	481.42	3.73	<.001	0.60-1.93
Gender=Male	0	0	-	-	-	-
Legal Status=Informal	-4.08	.78	582.65	-5.24	<.001	-5.60 - -2.55
Legal Status=Forensic	-3.54	.47	528.80	-7.56	<.001	-4.46 - -2.62
Legal Status= Civil	0	0	-	-	-	-

Table 5: Mean change in HoNOS-Secure item and total scores per assessment

Item	Estimate	Std. Error	95%CI	t	df	p	d
Scales 1 to 12							
1	-.09	.03	[-.15, -.02]	-2.72	166.49	.01	.42
2	-.02	.03	[-.08, .04]	-0.68	16533.88	.50	.01
3	-.01	.02	[-.04, .02]	-0.48	62181.46	.64	.00
4	.02	.03	[-.05, .09]	0.68	10.93	.51	.41
5	.02	.04	[-.05, .09]	0.63	12430.68	.53	.01
6	-.09	.03	[-.15, -.03]	-2.79	217.67	.01	.38
7	-.02	.03	[-.07, .04]	-0.54	123.40	.59	.10
8	-.05	.04	[-.13, .02]	-1.43	149.53	.16	.23
9	-.03	.03	[-.10, .04]	-0.80	125.37	.43	.14
10	-.06	.03	[-.12, .01]	-1.66	156.85	.10	.27
11	-.01	.02	[-.05, .04]	-0.26	55.47	.79	.07
12	-.06	.03	[-.11, -.01]	-2.28	20689.58	.02	.03
Total	-.37	.17	[-.71, -.04]	-2.20	118.96	.03	.40
Security scales							
A	-.07	.03	[-.14, -.01]	-2.18	105.11	.03	.42
B	-.04	.03	[-.11, .02]	-1.27	69.95	.21	.30
C	-.04	.02	[-.08, .00]	-2.14	277.95	.03	.26
D	-.08	.02	[-.12, -.04]	-3.64	171.83	**<.001**	.55
E	-.13	.03	[-.18, -.07]	-4.67	216.92	**<.001**	.63
F	-.05	.03	[-.10, .01]	-1.67	110.11	.10	.32
G	.00	.03	[-.06, .06]	0.14	60.18	.89	.04
Total	-.40	.11	[-.62, -.18]	-3.62	166.63	**<.001**	.56

N.B. **Bold** text indicates statistical significance at Bonferroni corrected P<.002 level. *d*=within subjects effect size.

Table 6: Number of people whose scores were in the normal and problematic range at the first assessment

Item	Normal score at first assessment - Final assessment outcome		Problematic score at first assessment - Final assessment outcome	
	Still in normal range (remained OK) n (%)	In problematic range (deteriorated) n (%)	In normal range (improved) n (%)	Still in problematic range (unchanged) n (%)
Scales 1 to 12				
1	220 (52.6)	51 (12.2)	77 (18.4)	70 (16.7)
2	345 (82.5)	17 (4.1)	39 (9.3)	17 (4.1)
3	407 (97.4)	2 (0.5)	8 (1.9)	1 (0.2)
4	271 (64.8)	46 (11.0)	45 (10.8)	56 (13.4)
5	245 (58.6)	54 (12.9)	57 (13.6)	62 (14.8)
6	203 (48.6)	32 (7.7)	58 (13.9)	125 (29.9)
7	287 (68.7)	35 (8.4)	61 (14.6)	35 (8.4)
8	80 (19.1)	67 (16.0)	69 (16.5)	202 (48.3)
9	73 (17.5)	42 (10.0)	51 (12.2)	252 (60.3)
10	152 (36.4)	47 (11.2)	65 (15.6)	154 (36.8)
11	389 (93.1)	11 (2.6)	16 (3.8)	2 (0.5)
12	372 (89.0)	12 (2.9)	29 (6.9)	5 (1.2)
Security scales				
A	13 (3.1)	11 (2.6)	33 (7.9)	361 (86.4)
B	52 (12.4)	27 (6.5)	45 (10.8)	294 (70.3)
C	2 (0.5)	1 (0.2)	22 (5.3)	393 (94.0)
D	0 (0.0)	1 (0.2)	2 (0.5)	415 (99.3)
E	30 (7.2)	5 (1.2)	54 (12.9)	329 (78.7)
F	13 (3.1)	23 (5.5)	26 (6.2)	356 (85.2)
G	0 (0.0)	1 (0.2)	1 (0.2)	416 (99.6)