Access to Electronic Thesis

Author:	Tina Perry
Thesis title:	Comparing self and clinician ratings on measures of mental health: A review of the literature and an assessment of the feasibility and utility of using the CORE-OM (Clinical Outcomes in Routine Evaluation – Outcome Measure) in secure hospitals
Qualification:	DClinPsy
Date awarded:	01 November 2010

Comparing self and clinician ratings on measures of mental health:

A review of the literature and an assessment of the feasibility and utility of using the CORE-OM (Clinical Outcomes in Routine Evaluation – Outcome Measure) in secure hospitals

Thesis submitted for the degree of Doctor of Clinical Psychology to the

University of Sheffield

Tina Jane Perry

BSc (Hons.), M.A, MSc.

July 2010

Declaration

The contents of this thesis have not been submitted to any other institution or for any other qualification.

Structure and Word Count

This thesis comprises two parts: (1) A literature review written in the format recommended by the journal *Psychology and Psychotherapy*: *Theory, Research and Practice* and (2) A research report written in the format recommended by *The Journal of Mental Health*. Guidance for both these journals is presented in Appendix 1.

Literature Review: A review of the associations between self-reported and clinician-reported measures of mental health for people presenting with a diagnosis of schizophrenia

Word Count (Excluding References): 7,730

Word Count (Including References): 9,339

Research Report: '**Determining the feasibility and utility of using the Clinical Outcomes in Routine Evaluation-Outcome Measure (CORE-OM) in secure hospitals'.**

Word Count (Excluding References): 11, 637

Word Count (Including References): 13,528

Total before appendices = 22,867

Total with appendices = 29,670

Acknowledgements

I am grateful to all those who participated in, or supported, this research including patients, nursing staff, ward managers, hospital managers, responsible clinicians and clinical directors. I would like to thank those closest to me, especially my partner for his never ending support and encouragement.

Mostly, I am eternally grateful for the support, guidance, wisdom and humour provided from both Prof. Barkham and Prof. Evans – it would have been so much harder without you both.

Overall Abstract

Literature Review: The objective was to investigate the association between self-reported and clinician-reported measures of anxiety and depression for people with a diagnosis of schizophrenia. ***Methods***: Electronic searches were conducted via Web of Knowledge, OvidSP, and CINAHL, analysis was conducted on 10 papers. ***Results***: All studies reported significant correlations between patient-rated and clinician-rated measures, although the strength of the correlations varied. Correlations were strongest when completed by researchers using appropriate measures. ***Conclusion***: The review profiles the association between patient and clinician-reported measures and suggests that self-report measures can be used meaningfully with patients with a diagnosis of schizophrenia.

Research Report: Government policy calls for the use of patient-reported outcome measures particularly in forensic services. ***Aims:*** This study aimed to (i) investigate the feasibility of the CORE-OM in secure hospitals, (ii) assess correlations between patient and nurse-ratings, and (iii) investigate patient responding style and the influence of insight on self-reported scores. ***Method*****:** Patients completed the CORE-OM and measures of insight and social desirability (SD). Nurses completed CORE-OM (staff version) and a functioning scale. Both participated in semi-structured interviews ***Results:*** Mean scores reported by patients and nurses were lower than those previously reported in prison, clinical and non-clinical samples, although the results do not show that patient scores were influenced by insight or SD. One significant correlation between the 2 perspectives was obtained was for the *functioning* domain. ***Conclusion*****:** Although the opinions of patients and nurses, and psychometric data suggest that the CORE-OM is acceptable and feasible within secure hospitals, the low scores are incongruent and there is the need for further research to understand this phenomenon.

Contents

Literature Review: A review of the associations between self-reported and clinician-reported measures of mental health for people presenting with a diagnosis of schizophrenia

Contents

Research Report: Determining the feasibility and utility of using the Clinical Outcomes in Routine Evaluation-Outcome Measure (CORE-OM) in secure hospitals.

Contents

List of Tables and Figures

Literature Review:

Research Report:

List of Appendices

Section One: Overall Appendices

Section Two: Literature Review Appendices

Section Three: Research Report Appendices

Title: A review of the associations between self-reported and clinician-reported measures of mental health for people presenting with a diagnosis of schizophrenia

1. Abstract

Purpose: The objective of this review was to investigate the association between self-reported and clinician-reported measures of anxiety and depression for people presenting with a diagnosis of schizophrenia.

Methods: Electronic searches were conducted via Web of Knowledge, OvidSP, the Cochrane Library and CINAHL along with visual scanning of reference lists and citation searches. Inclusion criteria were: (1) peer reviewed empirical study, (2) conducted and published in English, (3) conducted with adults with a diagnosis of schizophrenia, and (4) statistical comparisons between patient- and clinician-rated measures of anxiety and depression. Included papers were reviewed for methodological quality using the Quality Checklist (Downs & Black, 1998) and the COSMIN checklist (Mokkink et al., 2010). A narrative analysis was conducted on 10 papers.

Results: All studies reported significant correlations between patient-rated and clinician-rated measures, although the strength of the correlations varied from .31 to .94. Although no significant differences were shown during the secondary analysis the means suggest that correlations were strongest when completed by researchers (rather than clinicians) using appropriate measures.

Conclusions: The review profiles the association between patient and clinician-reported measures and suggests that self-report measures can be used meaningfully with patients presenting with a diagnosis of schizophrenia in clinical and research settings.

2. Introduction

2.1. Background

Recent Government policy has reinforced the importance of involving patients in their care (Department of Health, 2001; 2006; 2008) with a move towards shared assessment (Chadderton, 1995). This continues to be relevant for the new UK coalition government with the publication of a White Paper stating the intention to expand the use of Patient Reported Outcome Measures (PROMS) throughout the NHS by April 2011 (*Equality and excellence: Liberating the NHS*, 2010). According to this White Paper, PROMS are a way of increasing shared decision making between patients and clinicians, leading to improved health outcomes (Fremont et al., 2001), improved satisfaction with services (National Endowment for Science, Technology and the Arts: [NESTA] 2010), and reduced cost (Wanless, 2002). However, currently many PROMS in mental health relate to the concept of quality of life (Schmidt, Garratt, & Fitzpatrick, 2000).

National guidelines for schizophrenia note the need to routinely monitor for co-existing conditions such as depression or anxiety (NICE, 2009). Co-morbidity rates for a range of disorders have been reported as follows: obsessive compulsive disorders, 12.1%; social phobia, 14.9%; generalised anxiety disorder, 0.9%; panic disorder, 8%, post-traumatic stress disorders, 12.4% (Achim et al., 2009); depression 42–60% (Markou, 1996). However, it is thought that self reported difficulties for those people presenting with a diagnosis of schizophrenia might differ from the professional view (National Collaborating Centre for Mental Health [NCCMH], 2010). Thus, while government guidelines reinforce the need to obtain patient reports, questions have arisen as to whether an individual experiencing schizophrenia will have the awareness or ability to self-report their own difficulties (Kring, Kerr, Smith, & Neale, 1992; Loew & Rappin, 1994).

To date, research assessing the validity of self-report through comparisons with clinician-rated measures has focused on measures of need, functioning, and quality of life. The findings are mixed, with some studies reporting good reliability and validity for measures (e.g., Schmidt, Garratt & Fitzpatrick, 2000; Slade et al., 1995) and others reporting that validity is reliant on factors such as affective bias (e.g., Atkinson, Zibin & Chuang 1997; Bowie, 2006), insight (Doyle, 1999), compliance (Voruganti, Heslegrave, Awad & Seeman, 1998), and cognitive ability (Nishiyama Ozaki & Iwata, 2009). However, considering the level of co-morbidity of depression and anxiety for people presenting with schizophrenia, there remains a need for further research into the relationship between self- and clinician-reported measures (Niv, Cohen, Mintz, Ventura & Young, 2007; Paulhus & John, 1998).

2.2. Rationale

A brief review of the literature was conducted to determine whether a review in this area had previously been conducted. The NHS Centre for Reviews and Dissemination (NHS CRD, 2009) suggested a search of the Database of Abstracts of Reviews of Effects (DARE), the Cochrane Library, and the Evidence of Policy and Practice Information Centre (EPPI). This search revealed a single review, namely a systematic review within the Cochrane Library (Gilbody, House & Sheldon, 2008). This paper sought to establish the value of using outcome measures with patients with schizophrenia by reviewing randomised controlled trials which compared the management and outcome of those who used outcome measures with those who did not. However, this single review was not limited to measures of mental health and, seeking to review only randomised control trials, was unable to find any studies that met the criteria.

2.3 Objectives

The objective of the current review was to investigate the associations between self-reported and clinician-reported measures of mental health for people presenting with a diagnosis of schizophrenia. The concept of association was operatationalised as reported correlations.

3. Method

The review was informed by guidelines on conducting systematic reviews (NHS CRD, 2009). Although there has been a move away from using the term *schizophrenia* for people presenting with psychotic symptoms (NCCMH, 2010), it was used throughout this review as it is a key search term for reviewing the published literature.

3.1. Identification of studies

Studies were identified through electronic databases, visual scanning of reference lists, and citation searches. Electronic databases accessed through Web of Knowledge (WOK) were Web of Science (1900–2010), BIOSIS previews (1969–2010), and MEDLINE (1950-2010). Databases accessed through OvidSP were British Nursing Index and Archive (1985–May 2010), Ovid MEDLINE (1950–June Week 1 2010), and PsycINFO (1806–June Week 1 2010). In addition, CINAHL (Cumulative Index to Nursing and Allied Health Literature: 1990-2010) was also searched. The last search was dated 17th June 2010.

Electronic search terms for those conducted via OvidSP, WOK and CINAHL were: (1) Psychosis OR psychotic OR schizophrenia OR schizophrenic, (2) self-report, OR patient-report, OR client-report, OR consumer-report, OR self-rated, OR patient-rated, OR client-rated, OR consumer-rated, OR self-assessment, OR self-assessed and (3) valid OR validity. All search terms were searched ‘within abstracts’ or ‘In Topic’ with limits of ‘English Language only’. Once the appropriate full texts had been obtained visual scanning of reference lists and citation searches were carried out.

3.2. Study selection

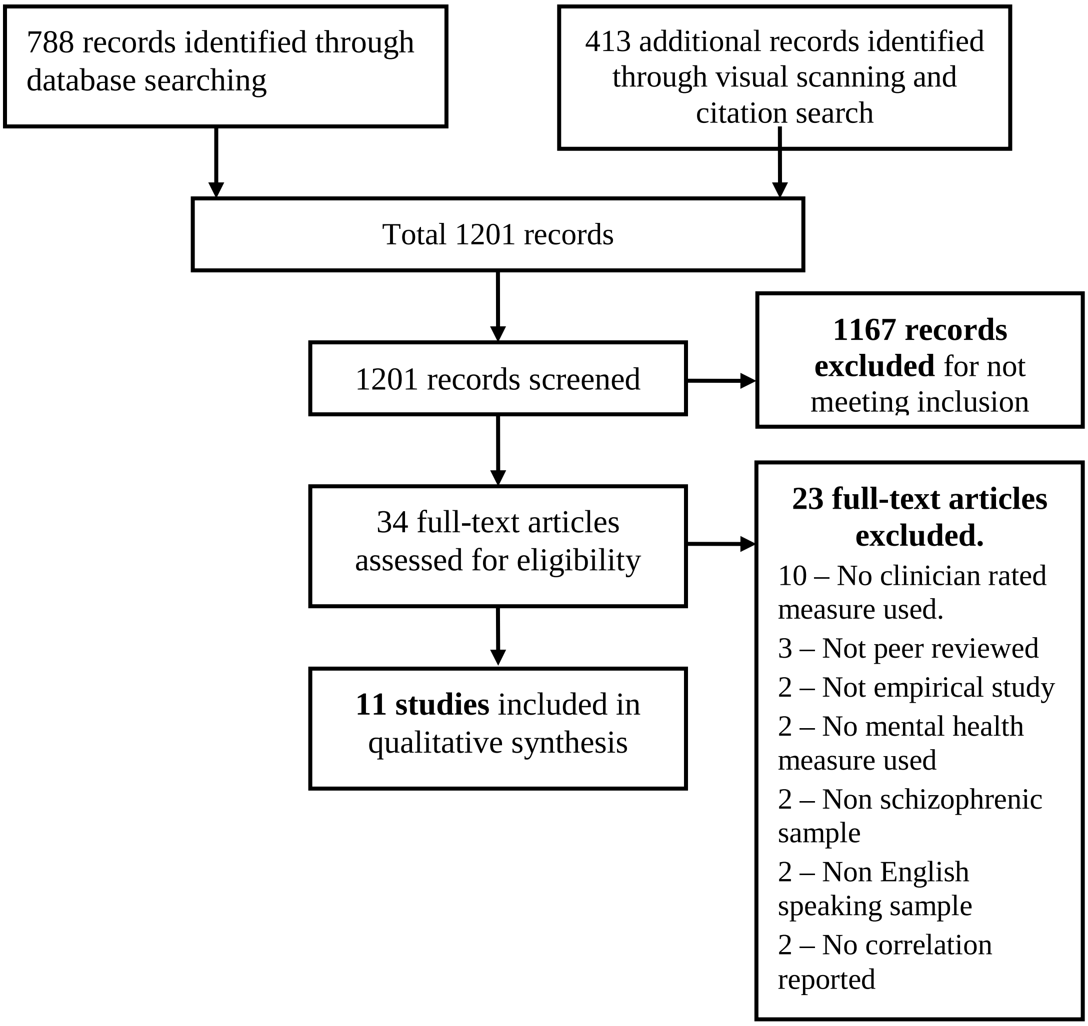
3.2.1 Eligibility criteria

To meet inclusion criteria, reviewed papers had to be based on an empirical study with adults presenting with a diagnosis of schizophrenia and be published in English in a peer reviewed journal. The research also had to report and statistically compare patient- and clinician- (or researcher-) rated measures of anxiety and/or depression. A review of alternative therapies has suggested that those who participate in research in non-English speaking countries are significantly more likely to report that a test intervention is superior to the control (Vickers, Goyal, Harland & Rees, 1998). From here they suggest that researchers conducting systematic reviews should consider whether studies published in non-English speaking countries are directly comparable to English speaking countries. As a result of this studies conducted in non-English speaking countries were excluded. Further exclusion criteria were studies that focused on medication or physical health and, for clarity, forensic, substance misuse or homeless populations.

3.2.2 Process for selecting studies

The search produced 1201 citations, 1167 were excluded leaving 34 full text articles which were inspected for exclusion criteria. From this a further 23 were excluded (see Figure 1 for details).

Figure 1: A CONSORT diagram of number of records screened, included, and excluded



3.2.3 Quality assessment

As recommended by the Centre for Reviews and Dissemination (2009), both methodological and reporting qualities were assessed. Downs and Black's Checklist for measuring study quality was employed (Downs & Black, 1998). A mean score is provided to help determine the quality of a paper. For the 27-item checklist the mean is 11.7 and is used as the cut-off below which papers are deemed to be poorer quality (Downs & Black, 1998). As the current review did not focus on interventions, 10 items focusing on interventions were omitted. Using the mean score as a percentage of the total value (43.3%), a new *average* score was calculated at 7.36.

To compensate for the items removed from the checklist, 8 items were included from the Consensus-based Standards for the selection of Health Measurement Instruments (COSMIN) checklist. This checklist can be adapted according to the needs of the reviewer (Mokkink et al., 2010). Thus, items included from the checklist were content validity, hypothesis testing, convergent validity, criterion validity, interpretability and generalisability. Methodological quality is rated on an ordinal scale of *excellent*, *good* and *fair*. If a study had major flaws, or the sample size was small, it was rated as *poor*. The scores on both checklists were combined to determine the overall quality of the papers under review. Papers that received a score of less than 7.36 on the Down and Black (1998) checklist, or those rated as ‘poor’ on the COSMIN checklist, were excluded from further analysis.

4. Results

4.1 Studies excluded following quality assessment

Scoring for each of the 11 papers according to the quality criteria is shown in appendix 2.1. At this stage one paper (Craig & Van Patta, 1976) that employed a small ‘schizophrenia’ sample from a hospital population was excluded from further analysis. A power analysis (performed by the reviewer) showed the study to be underpowered when reporting results solely for the schizophrenic sample (r = .09, alpha = .05, power = .08). Hence, 10 studies were retained for review.

4.2 Overview of study characteristics for the reviewed studies

Table 1 reports data for the recommended categories of setting, sample, measures, and statistical outcomes and the total quality checklist score (Stock, 1994). Details concerning the design, statistical analysis and power are presented in Appendix 2.2. Three papers focussed exclusively on depression while five focussed on both anxiety and depression. The remaining two papers used measures that addressed anxiety and depression but did not report correlations specific to these areas and, as such, were reviewed further under the section entitled *General Mental Health*. The total sample size for patients presenting with schizophrenia in the 10 studies was 990.

4.3 Review

Papers are reviewed in order of their total score on the Quality Assessment and COSMIN Checklists (i.e., highest scored papers reviewed first) with the total possible score being 26. Papers that obtained the same score were reviewed in alphabetic order. With the mean quality score of 21.4, papers with a rating of less than 21 received less focus in the review. For clarity the results are presented under the headings of *Anxiety and Depression* and *General Mental Health.* A list of all acronyms for the measures discussed can be seen in appendix 2.3. All reported correlation co-efficients are presented in Table 1. The strength of correlations are reported as either weak (r = .10 - .29), moderate (r = .30 - .49) or strong (r = .50 – 1.0) according to the criteria of Cohen (1988). Correlations not related to anxiety or depression are not discussed.

Table 1: Study characteristics for the reviewed papers

Author, Location, Score	Sample	Measures	Correlation Co-efficient (r)
Addington **1993** **Canada** ***24***	**Total = 150** In-patient = 50, Outpatient = 100 Mean Age = 39.9 (in) 43.3 (out) 48% Male (in) 63% Male (out) Mean admissions = 6.3 (in) 5.2 (out)	CR – Calgary Depression Scale (CDS) SR – Beck Depression Inventory (BDI)	**Inpatients (CDS & BDI)** Assisted = .94 Not Assisted = .82 Total = .84 **Outpatients (CDS & BDI)** Assisted = .69 Not Assisted = .69 Total = .69
Lindenmayer **1992** **USA** ***23***	**Total = 26 (inpatients)** Mean Age = 33.15 (SD = 7.94) 92% Male Duration = 13.46 years (SD = 7.63)	CR- Hamilton Depression Scale (HDS) CR- Positive and Negative Syndrome Scale (PANSS) NR- Nurses Observation Scale for Inpatient Evaluation (NOSIE) SR- Psychiatric Symptom Index (PSI) SR- Personality Profile Index (PPI) SR- The Mood Scales (MS)	HDS & PSI = .58 HDS & MS = .58 HDS & PPI = .22 PANSS & MS = -.12 PANSS & PPI = .04 PANSS & PSI = -.18 NOSIE & PSI (depression) = .02 NOSIE & MD (sad) = -.15 NOSIE & PPI = .06

Morlan **1998** **USA** ***22***	**Total = 27 (outpatients)** Age = 24-76 years 67% Male 52% Caucasian, 22% African-American, 7% Native American, 7% Hispanic, 11% not reported.	CR – Brief Psychiatric Rating Scale (BPRS) SR = Brief Symptom Inventory (BSI)	BPRS &BSI Depression = .69 BPRS & BSI Anxiety = .46
Hunter **2004** **Scotland** ***22***	**Total = 695** (30% schizophrenia, 48% affective disorder, Anxiety disorder 4%, Other 18%) Mean Age = 42 (SD = 12.2) 43% Male Duration Mean = 7.2 years (SD = 8.8)	CR and SR – Health of the Nations Outcome Scale (HONOS) SR and CR – Avon Mental Health Measure (AVON)	CR HONOS & SR HONOS = .4[k] SR AVON & CR AVON = .52[k]

Blanchard **1992** **USA** ***21***	**Total = 30** Inpatients = 25, Outpatients = 5 Mean Age = not reported 46% Male.	CR –BPRS CR- Raskin Depression Scale (RDS) SR – Symptom Checklist-90 (SCL-90)	**Time 1: BPRS &SCL-90** Anxiety = -.10 Depression = .05 **Time 1: RDS & SCL-90** Anxiety = .14 Depression = .34 **Time 2: BPRS & SCL-90** Anxiety = .67 Depression = .80 **Time 2: RDS & SCL-90** Anxiety = .56 Depression = .72
Hamera **1996** **USA** ***21***	**Total = 29 (outpatients)** Mean Age = 38 (range = 22-65) 59% Male Duration = 16 years (SD = 10.6) 53% more than 5 admissions90% White, 7% Black, 3% Other	CR = BPRS SR - Symptom Intensity Questionnaire (SIQ) SR- Symptom Distress Questionnaire (SDQ)	BPRS & SIQ = .66 BPRS & SDQ = .81

Huppert **2002** **USA** ***21***	**Total = 33** outpatient Mean Age = 36.63 (SD = 9.28, range 19-52) 39% Male 61% Caucasian, 18% Black, 3% Latino, 3% Asian, 15% unanswered.	**CR** Anxiety Disorders Interview Schedule for DSM-IV (ADIS) subscale Social phobia (SP) Panic Disorder (PD) Obsessive Compulsive Disorder (OCD) Generalised Anxiety Disorder (GAD) Major Depression (MD) PANSS- Anxiety PANSS - Depression SR –BDI Depression, Anxiety, and Stress Scale (DASS D, A & S) Anxiety Sensitivity Index (ASI) Obsessive Compulsive Inventory (OCI) Penn State Worry Questionnaire (PSWQ) Social Interaction Anxiety Scale (SIAS) Social Phobia Scale (SPS)	**Depression** BDI & ADIS = .58 BDI & PANSS D = .55 DASS D and ADIS = .57 DASS D& PANSS D = .59 **Anxiety (all PANSS A)** PANSS & ASI = .27 PANSS & DASS A = .64 PANSS & DASS S = .50 PANSS & PSWQ = .33 PANSS & SIA = .24 PANSS & SPS = .24 ADIS PD & ASI = .18 ADIS PD & DASS = .41 ADIS GAD & DASS S = .66 ADIS GAD & PSWQ = .44 ADIS SP & SIA = .40 ADIS SP & SPS = .42

Preston **2003** **Australia** ***21***	**Total = 69** 60% Male (mean age = 24.6) 40% Female (mean age = 28.5) Recently Diagnosed.	CR – PANSS SR – BSI	PANSS & BSI Depression = .46 PANSS & BSI Anxiety = .48
Addington **1992** **Canada** ***20***	**Total = 150** In-patient = 50, Outpatient = 100 Mean Age = 43 59% Male, Mean admissions = 5.5	CR - CDS CR- HDS CR- BPRS SR- BDI	CDS & BDI = .79 HDS & BDI = .77 BPRS & BDI = .73
Bell **2007** **USA** ***19***	**Total = 273** Outpatients Mean Age = 43 (SD = 8.6) 86% Male 65% Single Number of Hospital Admissions = 10 (SD = 10.8)	CR –PANSS CR – Scale to Assess Unawareness of Mental Disorder (SUMD) SR – NEO Five Factor Inventory (NEO) SR – Eyesenck Personality Questionnaire (EPQ) SR – BDI	NEO Neuroticism & PANSS Anxiety = .37 EPQ Neuroticism & PANSS Anxiety = .31 No comparisons reported for BDI

1 Only measures relating to depression and anxiety reported in the table. 2. CR = Clinician Rated, SR = Patient Self Rated, NR = Nurse Rated. K-.Kappa co-efficient reported

4.3.1 Anxiety and depression

(i) Better quality studies (ratings above 21)

Addington, Addington and Maticka-Tyndale (1993): This study investigated whether the Beck Depression Inventory (BDI: Beck, Ward, Mendelson, Mock & Erbaugh, 1961) was associated with the researcher-rated Calgary Depression Scale (CDS: Addington, Addington, & Schissel, 1990). The results yield one of the strongest correlations between objective and self-reported measures within the reviewed papers. The authors reported that some patients needed assistance to complete the measures (34% of inpatients, 17% of outpatients) and correlations were stronger for those requiring assistance. They were also stronger for inpatients (compared to outpatients) although the authors do not provide an explanation for this result. Despite being rated as high in terms of methodological quality, there are limitations as the authors report disagreement between researchers on the diagnosis of some participants but further details are not provided.

Overall, Addington et al. (1993) report strong associations between self- and researcher-rated measures and suggest that the former are appropriate for use with people presenting with schizophrenia. However, the author’s note that a large proportion of the inpatient sample were unable to complete the measure unassisted. Therefore, the authors suggest that both self- and researcher-report are beneficial in the assessment of depression for those presenting with schizophrenia.

Lindenmayer, Kay and Plutchik (1992): These authors hypothesised that self-report measures for patients experiencing schizophrenia would show weak correlations with clinician ratings. Inpatient participants completed the Mood Scale (MS; Plutchik, 1989), the Psychiatric Symptom Inventory (PSI; Ilfeld, 1977), and the Personality Profile Inventory (PPI; Plutchik, 1989). Clinicians completed the Positive and Negative Syndrome Scale (PANSS; Kay Fiszbein, & Opler, 1987), and the Hamilton Rating Scale for Depression (HDS; Hamilton, 1960) while nurses completed the Nurses Observation Scale for Inpatient Evaluation (NOSIE; Honigfeld, Gillis, & Klett, 1966). No statistical differences were found between patients who did and did not take part, suggesting that the consenting sample was representative of the approached sample. Correlations between clinician and patient ratings varied from negative to strong according to the measure used. Correlations were strong when clinicians used the HDS but either negative or weak when using the PANSS. Furthermore, correlations between nurse- and patient-rated measures were very weak.

There are several limitations to this study. Firstly, the patient sample was resident on a specialist research unit and may not be representative of the general population of inpatients with schizophrenia. Secondly, the sample was predominantly male (92%) and, as no analysis is reported for differences between male and female participants, it is unclear whether the results fully represent female patients with schizophrenia. Thirdly, although a researcher was available to support patients, no data was reported as to how many patients required assistance. However considerable difficulties arise in the choice of measures used within the study. Firstly, although the PPI addressed depression, it is designed as a personality assessment and the low correlations reported suggest that it may be an inappropriate comparison measure. Secondly, the NOSIE (nurse-rated) focuses on behaviour and the authors acknowledge that nurses may be rating different aspects of depression than patients when using this measure, thereby accounting for the poor correlations. Thirdly, the PANSS rating of depression is based on a single question. It is likely that this explains the low correlation between patient self–report and clinician ratings on the PANSS as the HDS (a 21-item measure of depression) shows significantly higher correlations

While the association between patient and clinician report is variable in this study, it is possible that correlations would have been stronger if the measures used throughout were more suitable. Overall the authors recommend that, when assessing depression for those people presenting with schizophrenia, both patients and clinicians complete measures. A further recommendation would be to ensure that measures used for comparison are directly comparable in the topics they measure.

Morlan and Tan (1998): These authors predicted a significant correlation between the Brief Psychiatric Rating Scale (BPRS; Overall & Gorham, 1962) and the Brief Symptom Inventory (BSI; Derogatis, & Melisaratos, 1983). Correlations were computed between the subscales of depression (BSI) with depressed mood (BPRS), anxiety (BSI and BPRS) among others. Strong correlations were shown for depression and moderate correlations for anxiety.

The authors acknowledge that using volunteers may have resulted in the sample not being fully representative. Further, the authors do not fully explore why the correlation for anxiety is lower than that for depression. Although most clients (82%) had a diagnosis of either schizophrenia or schizoaffective disorder, 18% technically met the exclusion criteria for this review. This paper was retained in the review due to the higher percentage of participants with schizophrenia and because of the relatively small number of papers open to review. As results were not reported according to diagnosis it is not possible to know whether those without schizophrenia produced different results.

In summary, the authors report a moderate to strong association between patient- and clinical-reported measures of anxiety and depression and suggest that either can be used for assessment. However, they cite research on the questionable discriminant validity of the BSI and BPRS (Boulet & Boss, 1991; Lukeoff et al., 1986). Therefore they recommend that both scales should be used alongside behavioural observation, symptom specific measures, and diagnosis for a more in-depth assessment.

Blanchard, Mueser and Bellack (1992): The authors of this study aimed to assess the association between patient- and researcher-rated self-report measures of depression and anxiety. Patients completed self-report measures at admission to hospital (Time 1) and at 6-month follow-up (Time 2). Although only 63% of the sample took part in the follow-up, the authors reported no statistically significant differences between those patients who did and did not take part in the study. The anxiety and depression scales of the SCL-90 (Derogatis, Lipman, & Covi, 1973) were used as the self-report index while researcher reports were taken from the BPRS scales for anxiety-depression and the Raskin Depression Scale (Lipman, 1982). The Thought-Disturbance scale of the BPRS was used to determine whether increased psychotic symptoms at follow-up were predicted by patient or researcher reports.

Correlations between patient- and researcher-reported measures at Time 1 were weak to moderate and non-significant. Yet correlations at Time 2 were strongly correlated.

Regression analyses to investigate whether negative mood at Time 1 was predictive of thought disturbance at Time 2 showed that while self-reported mood accounted for 29% of the variance in Time 2 thought disorder scores, follow up correlations showed only anxiety to be significantly (negatively) correlated with thought disturbance ($r = -.53$). This suggests that higher levels of self-reported anxiety at admission are related to lower levels of psychotic symptoms at follow-up. The authors rejected several explanations accounting for the differences between correlations at Times 1 and 2 including problems with the initial researcher rating, and patient familiarity at Time 2. From the correlations at Time 2 it is assumed that the measures used were appropriate. To explain the results the authors note that patient scores on the thought disturbance scale were higher at Time 1 (male mean = 13.36, SD = 5.53; female mean = 13.20, SD = 5.74) than at Time 2 (male mean = 10.25, SD = 2.54; female mean = 10.53, SD = 5.60). They suggest this may have had an impact on patients' presentations or openness to questioning during interviews conducted for the researcher-rated reports.

The study has several limitations. Firstly, the sample used comprised those people who consented and who were currently stabilised. While analyses showed no differences between the original and follow-up sample in relation to age, education, race, hospitalisation and symptomology, it is possible that participants in the follow-up group were more likely to agree with mental health professionals as evidenced by their having remained in the mental health system and consenting at Time 2. Thus, there is a possibility that those participants who were not followed up may have yielded scores that continued to show poor correlations with researcher ratings. Despite this, the study provides interesting results and hypotheses for low correlations between patient- and clinician-report measures. The authors conclude that self-report measures are valid for use with patients presenting with schizophrenia and that they may, in fact, be a more accurate predictor of outcome than researcher rated measures.

Huppert, Smith and Apfledorf (2002): In this study the authors used a variety of self-report measures of depression and anxiety with a sample of outpatients experiencing schizophrenia who were not in the midst of a psychotic episode. Forty-six patients with a diagnosis of anxiety or depression (but not schizophrenia) were compared on the same measures. Self-report measures for depression were the BDI (Beck et al., 1961) and the Depression, Anxiety and Stress Scale (DASS; Brown Chorpita, Korotitisch, & Barlow, 1997). A range of self-report measures was used for a variety of anxiety disorders. These were:

(i) The Anxiety Sensitivity Index (ASI; Taylor, 1999) and DASS – anxiety (Brown et al., 1997) for panic disorder

(ii) The Obsessive Compulsive Inventory (OCI; Foa, Kozak, Salkovskis, Coles & Amir, 1998) for Obsessive Compulsive Disorder (OCD)

(iii) The Penn State Worry Questionnaire (PSWQ; Brown, Antony & Barlow, 1992) and DASS-stress, (Brown et al., 1997) for Generalised Anxiety Disorder (GAD)

(iv) The Social Interaction Anxiety Scale (SIAS; Mattick & Clarke, 1998) and Social Phobia Scale (SPS; Mattick & Clarke 1998) for social phobia

The Anxiety Disorders Interview Schedule for DSM-IV (ADIS; including a scale on depression for co-morbidity and differential diagnosis; Brown, DiNardo, Lehman, & Campbell, 2001) and the PANSS were used as researcher-rated comparators. Only scores on the PSWQ differed significantly between the anxiety and schizophrenia samples (t (76) = 2.97, p<.01) with the anxiety group reported higher scores. The results yielded correlations of various strengths. While all correlations between self- and research-rated measures of depression were strong, correlations for anxiety ranged from .18 –to .66.

While the weak correlations between the PANSS and the SIA and SPS are likely to be due to the PANSS's lack of in-depth assessment of anxiety (Huppert et al., 2002), the authors note that the weak correlations using the ASI and the PSQW were unexpected and unlikely to be due to measurement error as the measures showed good reliability in this sample. After partialling out scores on the BDI, the authors suggested that the non-significant findings were not due to over reporting of distress. Instead, they suggested that anxiety constructs might be different in those people presenting with schizophrenia. Comparisons of the mean ASI scores showed scores for the schizophrenia sample (mean = 32.27, SD = 16.55) to be notably (but not significantly) higher than for the anxiety sample (mean = 28.83, SD = 13.64). The authors suggested that those presenting with schizophrenia might be more sensitive to anxiety symptoms than those without schizophrenia.

The authors were aware of the limitations of the small sample especially when considering the number of analyses conducted. However, they explicitly acknowledged having tolerated Type I errors in order to begin addressing important questions about self-report. Further limitations are also apparent. Firstly, as previously noted, the use of the single item PANSS may have impacted on the reliability of the measure. Thus, it is possible that correlations would have been higher on a multi-item measure. Secondly, the authors suggested that most patients who were approached did not take part either because they were unable to do so or because they had limited time. However this appears to be speculative. Furthermore, it is also proposed that those currently experiencing a psychotic episode would be unable to complete the measures. However, this is not supported by further information and contrasts with the earlier findings of Addington et al. (1993) who showed that 60% of inpatients experiencing relapse could complete self-report measures. Thirdly, the sample in this group were left alone to complete a large battery of measures and it is possible that their ability would have increased if fewer measures had been used or if a researcher had been available to assist them. Finally, no reference is made to whether order effects were taken into account when administering these questionnaires. However, this paper is one of the few to attempt to explain differences between the strengths of correlations for anxiety and depression. In summary, the authors mostly report moderate to strong correlations between self- and researcher-reported measures, further suggesting that self-report measures are both reliable and valid for use with people presenting with schizophrenia.

Preston and Harrison (2003): These authors used the BSI and the PANSS to determine whether patients recently diagnosed with schizophrenia rated themselves similarly to clinicians on measures of mental health. This is one of the few papers to show little difference in the reported correlations for anxiety and depression, with both diagnoses showing moderate correlations. However, as an analysis of correlations was not the main focus of this paper, a level of detail is lost. Comparisons using the PANSS did not differentiate between different mental health needs with the *general symptoms* score including depression, somatisation, anxiety and phobic anxiety. This leads to the suggestion that correlations may be influenced by poor discriminate validity. As a result the authors recommend the need to conduct further research using the same measure for both patient and clinician.

While problems with generalisability may be suggested, this appears to be the only paper aimed to understand the associations between patient and clinician reports for recently diagnosed patients and suggests that even those who are new to mental health systems can successfully self-report their difficulties.

(ii) Studies below quality threshold (ratings below 21)

Addington, Addington, Maticka-Tyndale and Joyce (1992): The authors in this study described the design of the Calgary Depression Scale (prior to the 1993 paper reviewed above). The CDS, the HDS, and the BPRS were used as researcher-rated measures. Strong correlations between all clinician rated measures and the self reported BDI were obtained.

Bell, Fizdon, Richardson, Lysaker and Bryson (2007) investigated the influence of insight on self-report measures of depression (BDI) and anxiety (NEO Five Factor Inventory-neuroticism scale: Costa & McCrae, 1989; and the Eysenck Personality Questionnaire: Eysenck & Eysenck, 1975). The Scale to Assess Unawareness of Mental Disorder (SUMD; Amador et al., 1994) was used to assess Insight. Clinicians' ratings were assessed using the PANSS. Despite the use of the BDI in this research, a clinician-rated comparator of depression was not used. Correlations between clinician- and patient-rated measures for anxiety were moderate. Neither self-reported anxiety nor depression was significantly affected by insight.

4.3.1 General Mental Health

The two papers reported in this section both received quality ratings above 21. They are discussed in this section as they do not provide separate correlations for anxiety and depression.

Hunter, McLean, Peck, Pullen, Greenfield, McArthur, Quinn, Eaglesham, Hagen & Norrie (2004) published the sole paper investigating correlations between clinicians and patients using the same measure. The Health of the Nation Outcome Scale (HoNOS; Wing, Beevor, Curtis, Park, Hadden & Burns, 1998) was rated by key workers, while patients were asked whether each item was relevant to their current problem. Then, the Avon Mental Health Measure (AVON; Markovitz, 1996) was completed by both patients and their key workers. Agreement ratings between patients and clinicians were in the moderate range for both HoNOS and AVON. Furthermore, the large sample in a naturalistic setting showed the feasibility of using a self-report measure in clinical care in the UK. While this is the sole paper to directly compare patient and clinician agreement on the same measure, it does not report agreement levels for those with a sole diagnosis of schizophrenia, nor does it specify the kappa statistics for the mental health domains of each measure. As such, it is plausible that much of the agreement related to practical issues given that the authors report that highest HoNOS kappa levels were for physical disability (0.50).

Hamera, Schneider, Potocky, and Casebeer (1996) examined the relationship between self- and researcher-report while developing a self-report symptom questionnaire for patients with schizophrenia. The measure was taken from the items in the BPRS and adapted to focus on symptom distress (SD) and symptom intensity (SI).

The 10 items comprised positive symptoms and negative symptoms tapping general symptoms (depression, anxiety, guilt and hostility). Strong correlations were reported between the patient and clinician reported measures.

The authors acknowledge limitations with the small sample size and the lack of comparison between responders and non-responders. However, the results show an association between patient- and clinician-reported measures and suggest that those people experiencing schizophrenia are able to accurately report their distress levels relating to anxiety and depression.

4.4 Secondary analysis.

The reviewed papers show a considerable variation in the correlations between patient- and other-reported measures of mental health. To assist in explaining these differences, Table 2 shows the study characteristics alongside all correlations reported in this review (ranked from largest to smallest). Data for the studies with quality ratings under 21 are in *italics* in the table. Several potential hypotheses arose and were analysed using SPSS 16. Differences between correlations are analysed using Fisher r-to-z transformation (Fisher, 1915).The correlations reported by Hunter et al. (2004) were removed from statistical analysis as they report kappa rather than intraclass correlations. These are discussed under the headings of methodological quality, clinician versus researcher as objective reporters, patient characteristics, differences in measures, and differences between anxiety and depression.

i. Methodological quality

The lowest correlation between self- and other-related measures is reported for a paper with one of the highest quality ratings (Lindenmayer et al., 1992). As a group, however, there was no significant correlation between quality score and the strength of the correlation co-efficient ($r = .17$, $n = 45$, $p = .26$). Further inspection of Table 2 suggests that differences in methodological quality scores do not influence the strength of correlations.

ii. Clinician vs. research as objective reporters

The majority of papers reporting high correlations derive from researcher- rather than clinician-completed measures with the top 21% of correlations for depression and anxiety being exclusively researcher rated. Mean correlations for researchers ($r = .58$, $n = 30$) appear higher than those for clinicians ($r = .26$, $n = 15$) although this difference was not significant ($Z = 1.40$, $p = .13$). These differences may relate to a drive for high levels of inter-rater reliability by researchers. In fact, of the 5 papers using researcher ratings, 4 provided data for inter-rater reliability (ranging from .77 – .89; see Appendix 2.2). In contrast, none of the clinician-rated measures reported inter-rater reliability correlations, although one reported training clinicians and ensuring acceptable inter-rater correlations prior to commencing the study (Hunter et al., 2004).

Table 2: Study characteristics ranked by correlation co-efficient

First author (Year)	**Correlation (measures)**	**Quality Score**	**Clinician or researcher rated**	**Patient setting (n)**	**Mental health area of focus**
Addington (1993)	.94 (CDS & BDI)	24	Researcher	Inpatient-Assisted (50)	Depression
Addington (1993)	.84 (CDS & BDI)	24	Researcher	Inpatient- Not Assisted (50)	Depression
Hamera (1996)	.81 (BPRS & SDQ)	21	Researcher	Outpatients (29)	General Mental Health
Blanchard (1992)	.80 (BPRS & SCL-90)	21	Researcher	Mixed (30)	Depression
Addington (1992)	*.79 (CDS & BDI)*	*20*	*Researcher*	*Mixed (50 in 100 out)*	*Depression*
Addington (1992)	*.77 (HDS & BDI)*	*20*	*Researcher*	*Outpatient (100)*	*Depression*
Addington (1992)	*.73 (BPRS &BDI)*	*20*	*Researcher*	*Outpatient (100)*	*Depression*
Blanchard (1992)	.72 (RDS & SCL-90)	21	Researcher	Mixed (30)	Depression
Addington (1993)	.69 (CDS & BDI)	24	Researcher	Outpatient – Assisted (100)	Depression
Addington (1993)	.69 (CDS & BDI)	24	Researcher	Outpatient – Not Assisted (100)	Depression
Morlan (1998)	.69 (BPRS & BSI)	22	Clinician	Outpatient (27)	Depression
Blanchard (1992)	.67 (BPRS & SCL-90)	21	Researcher	Mixed (30)	Anxiety
Hamera (1996)	.66 (BPRS & SIQ)	21	Researcher	Outpatient (29)	General Mental Health
Huppert (2002)	.66 (ADIS GAD & DASS Stress)	21	Researcher	Outpatient (33)	Anxiety

Huppert (2002)	.64 (DASS & PANSS)	21	Researcher	Outpatient (33)	Anxiety
Huppert (2002)	.60 (DASS & ADIS)	21	Researcher	Outpatient (33)	Anxiety
Huppert (2002)	.59 (DASS & PANSS)	21	Researcher	Outpatient (33)	Depression
Huppert (2002)	.58 (ADIS & BDI)	21	Researcher	Outpatient (33)	Depression
Lindenmayer (1992)	.58 (HDS & PSI)	23	Clinician	Inpatient (26)	Depression
Lindenmayer (1992)	.58 (HDS & MD-sad)	23	Clinician	Inpatient (26)	Depression
Huppert (2002)	.57 (DASS & ADIS)	21	Researcher	Outpatient (33)	Depression
Blanchard (1992)	.56 (RDS & SCL-90)	21	Researcher	Mixed (30)	Anxiety
Huppert (2002)	.55 (PANSS & BDI)	21	Researcher	Outpatient (33)	Depression
Hunter (2004)	.52- Kappa (AVON & AVON)	22	Key Worker	Mixed (104 In, 591 Out)	General Measure
Huppert (2002)	.50 (PANSS & DASS Stress)	21	Researcher	Outpatient (33)	Anxiety

Preston (2003)	.48 (PANSS & BSI)	21	Clinician	Outpatient (69)	Anxiety
Preston (2003)	.46 (PANSS & BSI)	21	Clinician	Outpatient (69)	Depression
Morlan (1998)	.46 (BPRS & BSI)	22	Clinician	Outpatient (27)	Anxiety
Huppert (2002)	.44 (ADIS GAD & PSWQ)	21	Researcher	Outpatient (33)	Anxiety
Huppert (2002)	.42 (ADIS SP & SPS)	21	Researcher	Outpatient (33)	Anxiety
Huppert (2002)	.41 (ADIS PD & DASS)	21	Researcher	Outpatient (33)	Anxiety
Huppert (2002)	.40 (ADIS SP & SIA)	21	Researcher	Outpatient (33)	Anxiety
Hunter (2004)	.40- Kappa (HONOS & HONOS)	22	Key Worker	Mixed (104 In, 591 Out)	General Measure
Bell (2007)	*.37 (NEO-FFI & PANSS)*	*19*	*Clinician*	*Outpatient (273)*	*Anxiety*
Huppert (2002)	.33 (PANSS & PSWQ)	21	Researcher	Outpatient (33)	Anxiety
Bell (2007)	*.31 (EPQ & PANSS)*	*19*	*Clinician*	*Outpatient (273)*	*Anxiety*

Huppert (2002)	.27 (PANSS & ASI)	21	Researcher	Outpatient(33)	Anxiety
Huppert (2002)	.24 (PANSS & SIA)	21	Researcher	Outpatient (33)	Anxiety
Huppert (2002)	.24 (PANSS & SPS)	21	Researcher	Outpatient (33)	Anxiety
Lindenmayer (1992)	.22 (HDS & PPI)	23	Clinician	Inpatient (26)	Depression
Huppert (2002)	.18 (ADIS PD & ASI)	21	Researcher	Outpatient (33)	Anxiety
Lindenmayer (1992)	.06 (NOSIE & PPI)	23	Nurse	Inpatient (26)	Depression
Lindenmayer (1992)	.04 (PANSS & PPI)	23	Clinician	Inpatient (26)	Depression
Lindenmayer (1992)	.02 (NOSIE & PSI)	23	Nurse	Inpatient (26)	Depression
Lindenmayer (1992)	-.12 (PANSS & MS- sad)	23	Clinician	Inpatient (26)	Depression
Lindenmayer (1992)	- 0.15 (NOSIE & MS- sad)	23	Nurse	Inpatient (26)	Depression
Lindenmayer (1992)	-.18 PANSS & PSI	23	Clinician	Inpatient (26)	Depression

iii. Patient location

Statistical analyses were conducted grouping patients by inpatient or outpatient. Those correlations reported for studies using mixed samples (n =5) were excluded from this analysis. While mean correlations were higher in the outpatient sample (r =.51, n = 29) than the inpatient sample (r = .26, n = 11) this difference was not significant (Z = 0.73, p = .23). Furthermore, it should be noted that 9 of the 11 inpatient correlations were reported by Lindenmayer et al (1992) who acknowledge the use of inappropriate measures in their study. Thus, the reported means may not reflect an actual difference between samples.

iv. Differences in measures

It is possible that the differences in correlations relate to differences in the measures used. Three of the 10 papers use the other-rated PANSS (Bell et al., 2007; Huppert et al., 2002; Preston & Harrison, 2003); while 4 use the BPRS (Addington et al., 1992; Blanchard et al., 1992; Hamera et al., 1996; Morlan & Tan, 1998). For patients 3 use the BDI (Addington et al., 1992, 1993; Huppert et al., 2002), and 2 use the BSI (Morlan & Tan, 1998; Preston & Harrison, 2003). Comparisons showed that mean correlations using the BPRS (r = .69, n = 7) were higher than those using the PANSS (r =.34, n = 12) although this difference was not significant (Z = .82, p = .21). For patient measures correlations were higher for the BDI (r =.78, n = 7) than for the BSI (r =.54, n = 5) although, again, this was not significant (Z = .51, p = .31)

As the PANSS takes on average 40 minutes to complete it is possible that correlations derived from the PANSS were lower due to patient fatigue. Previous research has suggested that motivation and cognitive states can affect scores on general questionnaires (Cannell, Miller, & Oskenberg, 1981). Furthermore, as the PANSS is based on one item for depression and anxiety it may be less psychometrically robust (Lindenmayer et al., 1992).

The BSI has been reported to have poor discriminant validity (Boulet & Boss, 1991; Lukeoff et al., 1986). Thus, it is possible that the BDI is a more appropriate self-report measure than the BSI, and the BPRS is more appropriate than the PANSS. Finally, it can be seen that the lowest correlations derive from measures that the authors acknowledge may not be suitable (Bell et al., 2007; Huppert et al., 2002; Lindenmayer et al., 1992). Thus, it is likely that the measures used throughout the review may have led to discrepancies in correlations.

v. Differences between anxiety and depression

Analysis of Table 2 appears to show that correlations for measures of depression are amongst the highest. However, when correlations between the two groups (with 2 correlations from measures of general mental health removed) were analysed no significant differences were seen ($Z = .38$, $p = .35$) although the means show that correlations for depression ($r = .50$, $n = 24$) were slightly higher than those for anxiety ($r = .40$, $n = 19$). Notably, the weak correlations (less than .30) for depression are reported by Lindenmayer et al. (1992) and may occur due to the use of inappropriate measures (as acknowledged by the authors). It is worth noting that even when using the same sample and the same measures (e.g., Morlan & Tan, 1998), correlations remained lower for anxiety. However, few papers attempt to explain the difference between anxiety and depression correlations. Huppert et al. (2002) suggest that patients with schizophrenia may be more sensitive to anxiety symptoms leading to higher levels of reported anxiety.

Blanchard et al. (1992) note that thought disturbance is negatively correlated with self-reported anxiety, suggesting that those who report high levels of anxiety have lower psychotic symptoms. However, as these differences are not fully explained, there is the need for further research.

5. Discussion

5.1 Summary

The current review suggests self-reported measures of mental health significantly correlate with clinician/researcher rated measures, with only one of 10 papers reviewed reporting correlations under .24 (Lindenmayer et al., 1992). However, there is some variability in the correlations. Although no significant differences were shown, the secondary analysis suggests that correlations are stronger when conducted by researchers (outside of patients' care). The most appropriate measures appear to be the BPRS (clinician-rated) and the BDI. The variety of self-report measures used for anxiety means it is not possible to determine which is preferable. The variation in correlations suggests that it may still be beneficial to take the viewpoints of both clinician and patient into consideration when assessing mental health symptoms (explicitly recommended by Addington et al., 1993, Lindenmayer et al., 1992; Morlan & Tan, 1998) particularly as two papers report that some patients are unable to complete the self-report measures (Addington et al., 1993; Hunter et al., 2009)

5.2 Clinical implications and critique of the review

The findings suggest that self-reported measures are associated with clinician-rated reports when the measures are appropriate. The secondary analysis also suggests the importance of training clinicians to use measures given that associations were stronger for those studies that reported high levels of inter-rater reliability. Taken together, the review implies that self-report measures can be accurately completed by patients with schizophrenia and, as such, can be used meaningfully within clinical settings. The fact that correlations were not perfect implies that patient and clinician perspectives can differ and there is a need to include the patient perspective during assessment and intervention, and particularly when evaluating outcomes.

There are several limitations of this review. Whilst a relatively comprehensive search of the literature was conducted, it did not include contacting authors to acquire any unpublished data. This may mean that the reviewed data is subject to a publication bias. However, the terms used for the search criteria were broad, which increased the probability that relevant published studies were not missed (NHS CRD, 2009). Furthermore, the review was subject to several exclusion criteria meaning the results may not be generalisable to those in non-English speaking countries, those from forensic, substance misuse or homeless populations, nor to other forms of self-report measures (i.e., quality of life, personality measures).

The process for assessing methodological quality may also have limitations. With the Quality Checklist (Downs & Black, 1998) only allowing scores of 1 (adequate) or 0 (not adequate) it was difficult to rank meaningfully the quality of studies (where some may have excelled in certain areas). This is supported by the suggestion in secondary analyses that measures completed by researchers showed higher correlations than those for clinicians and it is possible the checklist would have benefited from additional criteria relating to inter-rater reliability. Furthermore, the checklist criteria state that a point should be given if the detail for the criteria are not reported but could be assumed to be adequate. This means that components of several papers were attributed a score (i.e., 1 additional point) when detail was lacking. The Quality and COSMIN checklists could only be used as a basis on which to assess the literature. By contrast, the review of papers provided a more refined critique, while the secondary analysis was necessary to explore issues informing clinical implications.

5.3. Further Research

While the current review begins to map out the relationship between self- and other-reported measures of mental health for those people presenting with schizophrenia, there is a clear need for further research in this area. For example, future reviews would benefit from incorporating data published in non-English speaking populations. Furthermore, this study focussed on symptoms of anxiety and depression and there is a need to review the relationship between self and other ratings on measures of psychotic symptoms. The measures used were varied thereby making it difficult to compare the reported correlations. Hence there is the need for further empirical research assessing the relationship between widely used self-report and clinician-reported measures (with the hope of a future meta-analysis). The review excluded studies with co-morbid substance use difficulties and those from homeless or forensic populations. Thus, there is the need to for further research to determine the association between self- and clinician reported measures in these populations. Furthermore, although recommended by Preston & Harrison, 2003 only one paper (Hunter et al., 2004) investigated the relationship between self- and clinician-ratings on the same measure leading to the need for further empirical work in this area.

6. Conclusion

In conclusion, this review suggests that self-report measures of mental health symptoms positively correlate with clinician/researcher methods although these correlations may be stronger when obtained in research rather than clinical settings. As the secondary analysis cannot provide conclusive explanations of the variability of correlations, it is suggested that both self- and other-rated measures are used within clinical settings in order to obtain a fuller understanding and further research is needed to clarify this relationship.

References:

Achim, A.M,, Maziade, M., Raymond, E., Olivier, D., Merette, C., & Roy, M.A.(2009). How prevalent are anxiety disorders in Schizophrenia? A meta-analysis and critical review on a significant association. *Schizophrenia Bulletin*. Accessed online http://schizophreniabulletin.oxfordjournals.org/cgi/content/abstract/sbp148v1

Addington, D., Addington, J., Schissel, B. (1990). A depression rating scale for schizophrenics. *Schizophrenia Research*. 3. 247-251.

Amador, X.F., Flaum, M., Andreasen, N.C., Strauss, D.H., Yale, S.A., Clark, S.C., & Gorman, J.M. (1994). Awareness of illness in schizophrenia and schizoaffective andmood disorders. *Archives of General Psychiatry, 51,* 826–836.

Atkinson, M., Zibin, S., & Chuang, H. (1997). Characterizing quality of life among patients with chronic mental illness: a critical examination of the self-report methodology. American Journal of Psychiatry, 154, 99– 105.

Beck, A.T., Ward, C.H.,Mendelson, M., Mock, J., & Erbaugh, J. (1961). An inventory for measuring depression. *Archives of General Psychiatry, 4,* 561–571.

Bell, M., Fizdon, J., Richardson, R., Lysaker, P., & Bryson, G., (2007) Are self-reports valid for schizophrenia patients with poor insight? Relationship of unawareness of illness ro psychological self-report instruments. *Psychiatry Research. 151*. 37-46

Blanchard, J., Mueser, K.T., Bellack, A.S. (1992) Self and interview-rated negative mood states in schizophrenia: Their convergence and prediction of thought disturbance. *Journal of Psychopathology and Behavioural Assessment. 14 (3)*. 277-290.

Boulet, J., & Boss, M.W. (1991). Reliability and validity of the Brief Symptom Inventory. *Psychological Assessment, 3,* 433–437.

Bowie, C.R., Twamley, E.W., Anderson, H., Halpern, B., Patterson, T.L., Harvey, P.D. (2006). Self-assessment of functional status in schizophrenia. *Journal of Psychiatric Research, 41*, 1012–1018.

Brown, T.A., Antony, M.M., & Barlow, D.H . (1992). Psychometric properties of the Penn State Worry Questionnaire in a clinical anxiety disorders sample. *Behavior Research and Therapy*, *30*, 33–37.

Brown, T.A., Chorpita, B.F., Korotitsch, W., & Barlow, D.H. (1997). Psychometric properties of the Depression Anxiety Stress Scales (DASS) in clinical samples. *Behavior Research and Therapy*, *35,* 79–89.

Brown, T. A., DiNardo, P. A., Lehman, C. L., & Campbell, L. A. (2001). Reliability of DSM-IV anxiety and mood disorders: Implications for the classification of emotional disorders. *Journal of Abnormal Psychology*, *110*, 49–58.

Cannell, C.F. Miller, P.V., & Oskenberg, L. (1981). Research on interviewing techniques. *Sociological Methodology, 11,* 389-437.

Chadderton, H. (1995). An analysis of the concept of participation within the context of health care planning. *Journal of Nursing Management, 3,* 221–228.

Cohen, J. (1988). Statistical power analysis for the behavioral sciences (2nd ed.). New Jersey: Lawrence Erlbaum.

Costa, P.T., & McCrae, R.R. (1989). *NEO Five-Factor Inventory*. Psychological Assessment Resources, Odessa FL.

Craig, T.J. & Van Patta (1976). Recognition of depressed affect in hospitalised psychiatric patients: staff and patient perceptions. *Diseases of the Nervous System, 37*, 561-566.

Department of Health (2001) *The expert patient: a new approach to chronic disease management for the 21st Century.* The Stationary Office: London

Department of Health (2006) *Essence of care: benchmarks for promoting health. The Stationary Office*: London

Department of Health. (2008). *High quality care for all: NHS next stage review - Final report*. The Stationary Office: London.

Department of Health. (2010). *Equity and excellence: Liberating the NHS*. The Stationary Office: London.

Derogatis, L.R., Lipman, R.S., & Covi, L. (1973). SCL-90: an outpatient psychiatric rating scale – preliminary report. *Psychopharmacology Bulletin, 9,* 13-28.

Downs, S.H, & Black, N. (1998). The feasibility of creating a checklist for the assessment of the methodological quality both of randomised and non-randomised studies of health care interventions. *Journal of Epidemological Community Health, 52,* 377-384.

Doyle, M., Flanagan, S., Browne, S., Clarke, M., Lydon, D., Larkin, E., O'Callaghan, E., (1999). Subjective and external assessments of quality of life in schizophrenia: relationship to insight. *Acta Psychiatrica Scandinavica, 99,* 466–472.

Fisher, R.A. (1915). Frequency distribution of the values of the correlation coefficient in samples of an indefinitely large population. *Biometrika, 10,* 507-521.

Foa, E.B., Kozak, M.J., Salkovskis, P.M., Coles, M.E., & Amir, N. (1998). The Validation of a new Obsessive-Compulsive Disorder Scale: The Obsessive-Compulsive Inventory. *Psychological Assessment*, *10,* 206–214.

Fremont, A.M., Clearly, P.D., Hargraves, J.L., Rowes, R.M, Jacobson, N.B. & Ayanian, J.(2001). Patient-centred processes of care and long-term outcomes of myocardial infarction. *Journal of General Internal Medicine, 16,* 800-808.

Gilbody, S., House, A., & Sheldon, T. (2003). Outcome measures and needs assessment tools for schizophrenia and related disorders. *Cochrane Database of Systematic Reviews* 2003, Issue 1. Art. No.: CD003081. DOI: 10.1002/14651858.CD003081.

Hamera, E.K., Schneider, J.K., Potocky, M., & Casebeer, M.A., (1996). Validity of self-administered symptom scales in clients with schizophrenia and schizoaffective disorders. *Schizophrenia Research, 19*, 213–219.

Hamilton, M. (1960). A rating scale for depression. *Journal of Neurology, Neurosurgery, and Psvchiatry, 23*, 56-62.

Honigfeld, G. Gillis, R.D, & Klett, C.J. (1966). NOSIE-30: A treatment sensitive ward behavior scale. *Psychological Reports, 19*, 180-l82.

Hopko, D.R., Averill, P.M., Cowan, K., & Shah, N. (2002). Self-reported symptoms and treatment outcome among non-offending involuntary inpatients. *Journal of Forensic Psychiatry & Psychology, 13*, 88-106.

Hunter, R., McClean,J., Peck, D., Pullen, I., Greenfield, A., Mcarthur, W., Quinn, C., Eaglesham, J., Hagen., & Norrie, J. (2004). The Scottish 700 outcomes study: A comparative evaluation of the Health of the Nation Outcome Scale (HoNOS), the Avon Mental Health Measure (AVON), and an Idiographic (OPUS) in adult mental. *Journal of Mental Health, 13,* 93-105.

Ilfeld, F.W., Jr. (1977). Current social stressors and symptoms of depression. *American Journal of Psychiatry, l34*, 161-166.

Kay, S.R., Fiszbein, A., & Opler, L.A. (1987). The Positive and Negative Syndrome Scale (PANSS) for schizophrenia. *Schizophrenia Bulletin, 13(2)* 261-275.

Kring, A.M., Kerr, S.L., Smith, D.A., & Neale, J.M. (1993). Flat affect in schizophrenia does not reflect diminished subjective experience of emotion. *Journal of Abnormal Psychology, 102*, 507–517

Lindemayer, J.P, Kay, S.R., & Plutchik, R. (1992). Multivantaged assessment of depression in schizophrenia. *Psychiatry Research, 42,* 199-207.

Lipman, R.S. (1982). Differentiating anxiety and depression in anxiety disorders: Use of rating scales. *Psychopharmacology Bulletin, 18,* 69-77.

Loew, F., & Rapin, H. (1994). The paradoxes of quality of life and its phenomenological approach. *Journal of Palliative Care*, *10*, 37–41.

Lukoff, D., Liberman, R.P., & Nuechterlein, K.H. (1986). Symptom monitoring in the rehabilitation of schizophrenic patients. *Schizophrenia Bulletin, 12*, 578–602.

Markou, P. (1996). Depression in schizophrenia: a descriptive study. *Australian and New Zealand Journal of Psychiatry*, *30,* 354-357.

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

Mattick, R.P., & Clarke, J.C. (1998). Development and validation of measures of social phobia scrutiny fear and social interaction anxiety. *Behavior Research and Therapy, 36,* 455–470.

Mokkink, L.B., Terwee, C.B., Knol, D.L., Stratford, P.W., Alonso, J., Patrick, D.L., Bouter, L.M., de Vet H.C.W. (2010). The COSMIN checklist for evaluating the methodological quality of studies on measurement properties: A clarification of its content. *Medical Research Methodology, 18*, 10-22.

Morlan, K.K., & Tan, S.Y (1998). Comparison of the Brief Rating Scale and the Symptom Inventory. *Journal of Clinical Psychology. 54,* 885-894.

National Collaborating Centre for Mental Health: (2010) *Schizophrenia: The National Guidelines on core interventions in the treatment and management of schizophrenia in adults in primary and secondary care – updated edition.* Accessed online at: http://www.nccmh.org.uk/downloads/Schizophrenia_update/Schizophrenia%20full%20guideline%20post-publication%20version.pdf

National Endowment for Science, Technology and the Arts: (2010*) The Human factor: How transforming healthcare to involve the public can save money and save lives'.* Accessed online at: http://www.nesta.org.uk/library/documents/the-human-factor.pdf

National Institute for Health and Clinical Excellence. guidelines (2009) *Schizophrenia: Core interventions in the treatment and management of schizophrenia in adults in primary and secondary care*. National Institute for Health and Clinical Excellence. London

Nishiyama, T, Ozaki, N, & Iwata, N. (2009). Use of questionnaire infeasibility in order to detect cognitive disorders: Example of the Centre for Epidemiologic Studies Depression Scale in psychiatry settings. Psychiatry and Clinical Neurosciences, 63**, 23-29.**

Niv, N., Cohen, A.N, Mintz, J., Ventura, J. & Young, A.S. (2007). The validity of using patient self report to assess psychotic symptoms in schizophrenia, Schizophrenia Research. *90*, 245-250.

Overall, J.E., & Gorham, D.R. (1962). The Brief Psychiatric Rating Scale. *Psychological Reports, 10,* 799-812.

Paulhus, D.L., & John, O.P. (1998). Egoistic and moralistic biases in self perception: the interplay of self-deceptive styles with basic traits and motives. *Journal of Personality, 66,* 1025–1060.

Plutchik, R. (1989). Measuring emotions and their derivatives. In R. Plutchik & H. Kellerman (Eds.), *The measurement of emotions*. New York: Academic Press.

Preston, N.J & Harrison, T.J (2003). The Brief Symptom Inventory and the Positive and Negative Syndrome Scale: Discriminant validity between a self-reported and observational measure of pathology. *Comprehensive Psychiatry, 44*, 220-226.

Schmidt, L.J, Garratt, A.M., Fitzpatrick, R. (2000) Child/parent-assessed population health outcome measures: a structured review. *Child care, health and development 28* (3) 227-238

Slade, M, Phelan, M Thornicroft, G & Parkman, S (1996) The Camberwell Assessment of Need (CAN): Comparison of assessment by staff and patients of the needs of the severely mentally ill. *Social Psychiatry and Psychiatric Epidemiology. 31* (3) 109-113

Stock, W.A (1994) *Systematic coding for research synthesis*. In H.Cooper & L.V Hedges (Eds.) The handbook of research synthesis. 125-138. New York Russell Sage Foundation.

Taylor, S. (1999). *Anxiety sensitivity.* Hillsdale, NJ: Lawrence Erlbaum Associated.

The NHS Centre for Reviews and Dissemination. (2009). Systematic reviews: CRD's guidance for undertaking reviews in health care. Centre for Reviews and Dissemination, University of York.

Vickers, A., Goyal, N., Harland, R., Rees, R. (1998). Do certain countries produce only positive results? A systematic review of controlled trials. *Control Clinical Trials, 19*, 159-166.

Voruganti, L, Heslegrave, A.G, Awad, G.A & Seeman.M.V (1998). Quality of life measurement in schizophrenia: reconciling the quest for subjectivity with the question of reliability. *Psychological Medicine, 28*, 165-172.

Wanless, D (2002) *Securing our future health: Taking a Long-Term View: Wanless report.* Department of Health. The Stationary Office. London.

Wing, J.K., Curtis, R.H. & Beevor, A.S. (1996). *HoNOS: Health of the Nation Outcome Scales: Report on Research and Development July 1993-December 1995*. London: Royal College of Psychiatrists

Research Report: ‘Determining the feasibility and utility of using the Clinical Outcomes in Routine Evaluation-Outcome Measure (CORE-OM) in secure hospitals’.

1. Abstract

Background: Government policy calls for the use of patient-reported outcome measures and previous research suggests that the patient perspective may differ from that of the clinician. This is relevant in forensic services, where current outcomes measures do not fully address the domains considered vital for outcome assessment. CORE-OM maps onto key domains identified in a recent review (Fitzpatrick et al., 2010). However, there is the need to determine whether it is appropriate for use in secure hospitals. **Aims:** This study aimed to (i) investigate the feasibility and utility of the CORE-OM in secure hospitals, (ii) assess the association between patient and nurse-ratings, and (iii) investigate patient responding style and the influence of insight on self-reported scores on the CORE-OM. **Method:** A mixed-methods approach was used. Patients from 3 secure hospitals completed the CORE-OM and measures of insight (ITAQ) and social desirability (MCSDS). Their key nurses completed a staff version of the CORE-OM assessing the patient and an assessment of functioning scale (MIRECC GAF). Both patients and nurses participated in a semi-structured interview to obtain opinions on using the CORE-OM in secure hospitals **Results:** Mean clinical scores reported by both patient and nurses were lower than those previously reported in prison, clinical and non-clinical samples, although the results do not show evidence that patient scores were influenced by insight or social desirability. There were no significant differences between patient and nurse reported scores and the only significant correlation between the two perspectives obtained for the *functioning* domain. Themes from the interviews related to *acceptability, feasibility, relevance, suitability, changes to treatment* and *understanding.* **Conclusion:** Although the opinions of patients and nurses, and the psychometric data obtained, suggest that the CORE-OM is acceptable and feasible within secure hospitals, the low scores are incongruent and there is the need for further research to understand this phenomenon.

2. Introduction

2.1. Background

With the advent of evidenced-based medicine, there has been a growing call for the adoption of outcome measures in clinical practice. This has been apparent in the US (e.g., Sederer, Dickey & Eisen, 1997) and also in the UK with the advent of the Department of Health's Outcomes Reference Group (Fonagy, 2004). One conclusion arising out of the work of this latter group was that it was essential UK NHS trusts used outcome measures to inform service delivery and that these should be anchored against the practitioner-completed Health of the Nation Outcome Scales (HoNOS; Wing et al., 1998). Callaly and Halleborne (2001) supported the importance of collaborating with patients when using outcome measures and recent government policy has emphasised the importance of the patient voice in service delivery (Bower, Gilbody & Barkham, 2006). The combination of outcome measures and the patient voice has culminated in the recent Department of Health espousing a plan for the use of patient reported outcome measures (PROMs) across mental health settings (see *High quality care for all*, 2008) and throughout the entire NHS by April 2011 (*Equity and excellence: Liberating the NHS*, 2010). Furthermore, PROMS has been shown to improve therapeutic alliance, outcomes, and satisfaction with care while reducing violent behaviour (Beauford, McNeil, & Binder 1997; Eisen, Dickey, & Sederer, 2000; Horvath & Symonds, 1991).

2.2. Comparing patient and clinician reported measures

The Department of Health's oscillation between first prioritising the clinician and then patient perspectives raises the question as to the relationship between these two stakeholder views. Although previous research has shown moderate correlations between patient and clinician measures of mental health for those with schizophrenia (see accompanying literature review), research in other areas has shown some discrepancy. When assessing concordance between patient and clinician ratings of need, agreement is often stronger for tangible areas of need relating to daily functioning (Najim & McCrone, 2005; Slade 1994). It is thought that these differences occur due to staff being more likely to observe difficulties in areas that they consider to be their responsibility (Najim & McCrone, 2005). Of the research investigating agreement between psychiatric in-patients and their nurses there appears to be relatively poor overall agreement (Farrell, 1991; Lindenmayer et al., 1992; Nolan, 1989). Furthermore, in forensic services an individual's needs may be lost in the drive to manage risk (Swinton, 2000). Although research investigating the use of outcome measures with forensic participants is lacking (Fitzpatrick et al., 2010), involuntary patients endorse fewer symptoms than either their clinician or a comparator voluntary patient group (Hopko et al., 2002). From this finding it is possible to predict that forensic mental health patients, who are involuntary patients, may report scores on outcome measures that are different to their clinicians (Taylor, 1998).

2.3 Forensic mental health

Secure hospitals within the UK aim to provide care to patients with severe mental health needs whilst maintaining the safety of the public (Health Select Committee, 2000). Patients within these settings have complex needs and experience a range of difficulties. In high secure settings 75% of men and 50% of women have a diagnosis of schizophrenia (Taylor, 1998). In relation to offending, 90% of men and 75% of women have committed a serious criminal offence and all are compulsorily detained with an average stay of 7.5 years (Taylor, 1998). In medium secure services 75% have a diagnosis of schizophrenia (Murray, 1996) while 14% meet diagnostic criteria for an affective disorder (Smith et al., 1991). In low secure settings schizophrenia is, again, the most common diagnosis, with levels of affective disorders higher than in other populations (Cripps et al., 1995; Mcclintock & Evans, 1995). Considering these difficulties, the need for accurate assessment of difficulties and outcomes is undeniable. However, there is a lack of consensus on appropriate outcome measures in forensic mental health settings (Cohen & Eastman, 1997; Fitzpatrick et al., 2010).

2.4 Outcome measures in forensic mental health

In a review of 308 studies using outcome measures in forensic mental health research Fitzpatrick et al. (2010) acknowledged that forensic services were one of the few who were obliged to provide evidence of outcomes for both client care and public safety. In relation to client care, a consensus group comprising 10 forensic mental health professionals identified 21 domains that they considered vital for outcome assessment in forensic mental health. Of these 21, 11 related to patients' mental health with the remaining items relating to offending, treatment variables, and physical health. While the review noted that the most common measures were for substance use (n=73), recidivism (n=65), and mental state (n=65), there were several problems with the measures for mental state. The review indicated that the most often used mental states measures in UK forensic services are the Symptom Checklist 90-Revised (SCL-90-R; Derogatis, Lipman &, Covi, 1973), the Beck Depression Inventory (BDI; Beck, Ward, Mendelson, Mock, & Erbaugh, 1961) and the Brief Psychiatric Rating Scale (BPRS; Overall & Gorham, 1962). While these are popular measures, the consensus group reported that the previously identified domains were not fully represented by these measures. Furthermore, research on the relevance and feasibility of using these measures in forensic mental health services is lacking. While it is acknowledged that research in forensic settings can be problematic the need to explore the use of mental health outcome measures remains (Cure, Chua, Duggan & Adams, 2005; Fitzpatrick et al., 2010). Research suggests that forensic outcome measures should assess the four domains of clinical symptoms, functioning, wellbeing, and risk to public safety (Atkisson, Cook, & Karno, 1992; Fitzpatrick et al., 2010), and with the current most commonly used outcome measures in forensic services relating to public safety, there is a need to investigate outcome measures that address the remaining three domains (Cohen & Eastman., 2000; Fitzpatrick et al., 2010)

2.5 The Clinical Outcomes in Routine Evaluation – Outcome Measure (CORE-OM; Barkham et al., 2001; Evans et al., 2002)

One measure that maps directly onto the four domains identified by Fitzpatrick et al. (2010) is the Clinical Outcomes in Routine Evaluation-Outcome Measure (CORE-OM; Barkham et al., 2001; Evans et al., 2002). The CORE-OM is a widely used self-report outcome measure in primary care psychological therapy and counselling services (Office of Health Economics, 2008) and assesses self-reported difficulties with wellbeing, psychological symptoms, functioning, and risk. Furthermore, the CORE-OM shows correlations with HoNOS of .50 and is suggested as a self-report alternative when clinicians are not trained to use the HoNOS (Leach et al., 2005). The Office of Health Economics report (2008) on NHS outcomes, performance and productivity recommended the use CORE-OM in all community mental health services. However, to date there have been few studies that have investigated the usefulness of the CORE-OM outside of primary care settings. The few exceptions have suggested that the CORE-OM is valid in those people presenting with borderline personality disorder (Whewell & Bonanno, 2000), those in secondary care (Barkham et al., 2005, Leach et al., 2005) and those in prison settings (McCloskey, 2001). The finding that the CORE-OM is suitable for offenders outside of mental health services leads to the suggestion that it may be feasible for use within a forensic mental health setting and McCloskey (2001) recommends the need for further research to determine the usefulness of the CORE-OM in secure hospital settings.

2.6 Determining the usefulness of an outcome measure

If outcome measures are to be used within forensic mental health settings they must be feasible and relevant to the client group (Fitzpatrick et al., 2010). According to Evans, Greenhalgh, and Connelly (2000) there is considerable variability between measures and it is vital that the chosen tool is fit for purpose. They provide a useful ‘mental health assessment critical appraisal checklist’ to assist clinicians in choosing the most appropriate measure for a client group and suggest four evaluative areas as follows:

1. Psychometric exploration – Is the measure reliable? Is there evidence of its validity? Is there evidence that it is responsive to change?
2. Stakeholder perspectives – Does the measure capture the views of multiple stakeholders (client, clinician)? Does it capture the views of those people at all stages/severity of a condition/subgroup?
3. Feasibility – Is it feasible to use this measure with this population? Is training needed? How long does it take to complete and analyse?
4. Utility – Is the measure acceptable and perceived as relevant by clients and clinicians? Does it provide additional information that is not already available? Will the information aid care planning and decision making?

2.7 Assessing validity in forensic settings

Assessing the validity of a self-report measure in forensic services is complex and potential difficulties with the responding style of forensic patients have been well documented (Heilburn, 1992; Rogers 1984, 1988). Responding styles in forensic inpatients include *accurate*, *malingering*, *defensive* or *irrelevant* (Heilbrun, Bennet, White, & Kelly, 1990). In forensic inpatients malingering is seen as the need to exaggerate symptoms in order to avoid a prison transfer or to gain access to additional medication or support (Hawes & Boccaccini, 2009; Helmes, 2009). Although clinicians have raised concerns about this response style (Jacobs, 2007), it appears to be a rare with only 9% of forensic inpatients responding in this way (Heilbrun, Bennet, White, & Kelly, 1990). In contrast, a defensive response style is more common with a reported rate of 22% (Heilbrun, Bennet, White, & Kelly, 1990). This involves patients minimising their difficulties either as a result of lack of insight (Selten, Wiersma, & Van den Bosch, 2000) or social desirability (Baer & Miller, 2002). Finally, an irrelevant responding style occurs when the patient is not engaged in the assessment process demonstrated through poor internal consistency on measures (Heilburn et al., 1990). To determine whether a proposed measure is valid, there is a clear need to ensure that self-reports are not significantly influenced by patient responding style. However, there is also the need to investigate the responses of staff. As previously seen, nurses may over-report difficulties (Farrell, 1991) or focus too much on the tangible difficulties (Najim & McCrone, 2005; Slade 1994). In order to determine differences between staff and patient ratings on an outcome measure, it is preferable that the same measure is used for both participants (Hunter, 2004; Preston & Harrison, 2003).

2.8 Aims

In light of the above background, this study aimed to investigate the acceptability, feasibility and utility of using the CORE-OM with patients in secure hospitals. Furthermore, it aimed to assess agreement on patient difficulties between patients and staff by comparing patient scores on the CORE-OM with a version of the same measure completed by their key nurse. There are four primary aims and three secondary aims. The initial aims are:

1. To establish basic psychometric data for both the patient and nurse completed CORE-OM in secure settings.
2. To determine whether there is a positive association between patient and nurse responses on each domain of the CORE-OM.
3. To test whether nurse ratings of patient functioning (on a global assessment of functioning scale) is associated with severity of difficulties as rated on CORE-OM, thereby providing evidence of internal reliability of response.
4. To establish staff and patient opinion on the acceptability, feasibility and utility of using the CORE-OM in secure hospitals using a brief semi-structured interview.

The secondary aims are:

5. To determine whether patients report significantly higher CORE-OM scores than their key workers (indicating a malingering response style).
6. To determine whether patient reported scores on the CORE-OM are related to social desirability (indicating a defensive response style).
7. To determine whether patient reported scores on the CORE-OM are related to level of insight.

3. Method

3.1. Design

The design employed a mixed methods approach although is predominantly quantitative but complemented with qualitative data obtained through the use of a semi-structured interview carried out after the administration of the CORE-OM.

3.2 Settings

The study was carried out in three secure hospitals in the United Kingdom. These comprised (1) the Mental Health Directorate of a high secure hospital provided by Nottinghamshire Healthcare NHS Trust, and (2) a medium secure (Alpha) and (3) a low secure (Affinity Healthcare) hospital both of which were located in Greater Manchester. The high secure hospital is one of three within the UK. There are approximately 100 patients in the Mental Health Directorate with twelve beds dedicated to deaf patients. Alpha hospital provides a medium secure mental health service to 40 male and 20 female patients (with 12 deaf patients). Affinity healthcare provides a low secure service to patients and comprises two low secure female wards, and two low secure male wards.

3.3 Sample

Whilst original inclusion criteria (see below) suggested that both men and women would be included in this study, the response rate for women in the low and medium secure sample was low with only 9 women originally consenting to meet with the researcher. Of these 9 women, 7 refused either on first or second meeting with the primary researcher. Of the other two participants, one took 50 minutes to complete the

CORE-OM (where the average completion time for men was under 6 minutes), whilst the other omitted over one-third (12 of 34 items) on the CORE-OM (seemingly due to confusion as a result of voice hearing). Furthermore, time constraints made accessing the female high secure sample problematic as the Women's Service is a directorate independent to the Mental Health Service (where ethical approval had been obtained) and a meeting with the clinical director could not be arranged within the time limits of the study. As a result the decision was made to exclude women from the research project and the data obtained from the two who did take part was not included in the analysis.

3.3.1 Study inclusion and exclusion criteria

Participants were identified through discussion with their Responsible Clinician according to the inclusion and exclusion criteria. Inclusion criteria were: men with a diagnosis of a psychotic disorder currently residing in either the low, medium or high security hospital. Those with a primary diagnosis of psychosis with a co-morbid personality disorder were included in the study. Exclusion criteria were: men who did not speak English and those registered with the learning disability (LD), deaf, or personality disorder (PD) services. In order to avoid distress, those people who the Responsible Clinician deemed likely to be distressed by the study were excluded.

3.3.2. Patient characteristics

While a total of 39 men and 9 women initially consented to take part, 12 refused when approached by the researcher. The reasons for withdrawal were (1) withdrawal after the researcher confirmed that participation was not compulsory (42%), (2) withdrawal after confirmation that taking part would not be used as evidence for discharging the patient from hospital (42%), and (3) withdrawal due to an unwillingness to have the results feedback to their clinical team (16%). As previously discussed the two consenting women were removed from further analysis. As those who withdrew from the research had not provided consent to access their information it was not possible to make comparisons between those who did, and did not, take part in the research. Thus the total sample comprised 34 men, all with a diagnosis of Schizophrenia. With regard to ethnicity 88% were White British, 6% were White Irish, 3% were South Asian and 3% were African. Mean age was 41.2 years (SD = 8.7). Average length of current admission (in years) at each security level was low (mean = 4.6, SD = 2.91), medium (mean = 5.7, SD = 1.3) and high (mean = 7.9, SD = 5.3). There was no significant difference between patient age (F = .334 (2, 31) p = .718) or length of stay (F = 1.909 (2, 31), p = .165) according to level of security. Further results are analysed using the sample as a whole group (n = 34).

3.4 Measures

Hard copies of all measures are shown in the appendices.

3.4.1 Measurement development

Three measures were developed for the purpose of this research. Two of these were semi-structured interviews used to obtain qualitative feedback on the acceptability, feasibility and utility of using the CORE-OM in secure settings. The third measure was designed as a comparator to the CORE-OM (patient version). The three new measures are discussed below.

i. The Brief CORE-OM Interview – Patient Version (Perry, Barkham & Evans, 2010a)

A semi-structured interview was developed by the researchers comprising 9 open-ended questions. Questions for both the patient and staff version were derived from the criteria of Evans, Greenhalgh and Connelly (2000) on selecting appropriate mental health needs assessments (see *Section 2.6*). To answer the questions on utility the interview asked about the acceptability of questions (*Were there any questions which you did/did not like and why?*), and perceived relevance (*Do you think it would be useful for a member of your clinical team to use with you?*). To answer questions on feasibility the interview asks '*would you prefer to complete the questionnaire on your own or with a key worker'.* The question to assess whether it captures the view of patients at all stages of care was '*Do you think it would have been useful if a member of your clinical team had used this questionnaire with you at admission?* In order to fully explore the patient perspective two more open ended questions (*When we filled out this questionnaire, what did you think about it?* and *Do you have any other comments about the questionnaire?*) were included. Furthermore, the questions in the interview were also designed to address beliefs about responses to questionnaires and treatment (*Did you feel you were able to be open in your responses?* and *Did you worry that answering this questionnaire would affect your treatment or make any difference to how long you stay in hospital?*) and perceived self understanding (*Do you feel you have a good enough understanding of your feelings to answer the questions?*). This was in relation to the literature on defensiveness and insight in forensic patients (Heilbrun, Bennet, White & Kelly, 1990).

ii. The Brief CORE-OM Interview – Staff Version (Perry, Barkham & Evans, 2010b)

This semi-structured interview was developed by the researchers comprising 9 questions. Again these were based on the criteria laid out by Evans, Greenhalgh and Connelly (2000) To answer questions on utility nurses were asked about the acceptability (*Were any questions difficult to answer?)* relevance (*Do you think this measure would have any use across the service as a standard, Do you think it is relevant to your patient group?)* and the potential for changes to treatment (*Do you it would give you any information you did not already know and so make a difference to care planning?).* To answer questions on feasibility nurses were asked: *Do you think your patient could complete the questionnaire alone or with a key worker*? As this is one of the few studies using the same measure for patients and staff, the questionnaire also sought to obtain staff views on answering on behalf of the patient (*Did you feel you had understanding of your patient to answer the questions,* and *Did you feel under pressure to give the 'right' answer?*). To obtain nurse overall perspective two general questions were asked (*What was your overall opinion of using this questionnaire* and *Do you have any other comments?*).

iii. CORE-OM Staff Version (CORE-OM SV)

The CORE-OM staff version was developed mindful of recent literature suggesting the need for comparisons between clients and clinicians on the same measure (Hunter, Mclean, Peck et al., 2004; Preston & Harrison, 2003). It was closely modelled on the patient version of the CORE-OM (Barkham et al., 2001, Evans et al., 2002) with the key change being rewording the stem question from '*Over the last week I have*' to '*In my opinion, over the last week ________ has*' (with the blank space available for the client's name). The patient version (see *Section 3.4.2*) is a self-report measure comprising 34 items that address four domains: *subjective wellbeing, symptoms, functioning,* and *risk* (Evans et al., 2002). Each domain comprises specific clusters of items as follows: Symptoms comprises item clusters focusing on anxiety, depression, physical symptoms and trauma, while Functioning comprise general functioning, close relationships, and social relationships. Risk comprises risk to others, and risk to self. The Wellbeing scale comprises 4 items which all relate to subjective wellbeing. Each item is scored on a scale from 0-4, representing ‘Not at All’, ‘Only Occasionally’, ‘Sometimes’, ‘Often’ and ‘Most, or ‘All of the Time’. For most items increased frequency is anchored to a higher score (so that ‘Most, or All of the time’ leads to a score of 4). However, these items are reverse scored on 8 items where each item is indicative of a positive statement (e.g. ‘*I have felt optimistic about my future*’). Half of the items focus on low intensity (e.g. ‘*I have felt tense, anxious or nervous*’) and half on high intensity problems (e.g. ‘*I have felt panic or terror*’). Mean item scores are obtained by adding the sum of item scores and dividing by the number of completed items, resulting in a score between 0 to 4. Clinical scores are obtained by calculating the mean score and multiplying it by 10, resulting in a score ranging from 0 to 40. Clinical scores will be used and reported throughout this study. Reported categories for clinical scores (Barkham, Mellor-Clark, Connell & Cahill, 2006) are healthy (0-5), low level (6-9), mild (10-14), moderate (15-19), moderate-severe (20-24) and severe (over 25). Domains, symptoms, clusters and scoring in this staff version are identical to the patient version which is described in more detail below. Permission to develop this version was obtained from Professor Evans acting on behalf of the CORE System Trustees. This measure will be referred to as the CORE-OM (SV).

3.4.2 Published Patient Completed Measures

i. The CORE-OM (Evans et al., 2002):

Details for the CORE-OM patient version have been described above. Further research has shown internal consistency of .75 -.94 (Barkham et al., 2007) and 1 week test-retest reliabilities between .60-.91 (Evans et al., 2002). CORE-OM forms are considered valid if up to three items are omitted (Evans et al., 2002). As this version of the CORE-OM was not designed for use with patients with learning disabilities, those from learning disabilities directorates were excluded from the sample. Furthermore, as the CORE-OM is currently being translated for use with a deaf population those in Deaf services were also excluded. There is evidence that those who have English as a second language omit significantly more items when completing the CORE-OM (Evans et al., 2002) and as such those who did not speak English fluently were excluded. Throughout this research this measure will be referred to as the CORE-OM (PV).

ii. Insight and Treatment Attitudes Questionnaire (ITAQ; McEvoy et al., 1989)

The ITAQ was designed to measure awareness of mental health problems (or ‘illness’) and insight into the need for treatment in patients with psychosis. The ITAQ comprises three domains: ‘Do you have a mental illness?’, ‘Do you need treatment?’ and ‘Do you need medication? These domains are proposed for the past, present and future (e.g., ‘Will you, in the future, need continued treatment for mental health problems?’). The measure comprises 11 items designed to elicit open-ended responses that are scored according to a coding criteria provided with the measure with higher scores indicative of better insight. McEvoy et al. (1989) reported high inter-rater reliability (.82), high construct validity (.85), and a test-retest reliability of .70.

iii. The Reynolds’ Short Form C of the Marlowe-Crowne Social Desirability Scale (MCSDS; 1982)

The Reynolds’ Short Form C of the Marlowe-Crowne Social Desirability Scale (1982) was derived from the Marlowe-Crowne Social Desirability Scale (1960). The original version of this questionnaire comprised 33 items thought to be culturally sanctioned but unlikely to occur. The items are designed to reflect a propensity towards positive self-representation. More recently it has been used in research as a measure of defensiveness (Desheilds, Tait, Gfeller, & Chibnall., 1995; Mann & James, 1998). The Reynolds’

Short Form C of the Marlowe-Crowne Social Desirability Scale (1982) comprises 13 items taken directly from the original version. Participants respond either ‘True’ or ‘False’ to items such as ‘*No matter who I’m talking to I’m always a good listener*’ and ‘*I have never been irked when someone expressed ideas very different from my own*’. A score of 1 is obtained for items that are endorsed in a socially desirable manner. Reynolds’ Short-Form C has been shown to have internal consistency ranging from .62 to .76 and test-retest correlations of .74 (Ballard, 1992; Reynold, 1982). Furthermore, Andrews and Mayer (2003) normed Reynolds’ Short Form C in forensic samples and, showed the mean for forensic samples (7.61) to be above the mean for the general population (5.37). Thus, it was used instead of the original version to reduce time costs in administering the measures.

3.4.3. Published Staff Completed Measures

i. The MIRECC Version of the Global Assessment of Functioning Scale (Niv, Cohen, Sullivan, & Young., 2007)

The MIRECC GAF measures *occupational functioning; social functioning and symptomatic functioning* on three subscales each with a rating from 1-100 (with 100 indicating highest level of functioning). All MIRECC GAF scales are divided into 10 equal intervals and include criteria for scoring within each interval (e.g. ‘*Works consistently in sheltered work*’ or ‘*Frequent interpersonal conflicts or withdrawal but still able to maintain some meaningful interpersonal relationships*’). Relationships with professional caregivers and functional problems due to physical or environment limitations should not be rated. Niv et al. (2007) showed the measure to have good reliability (0.99) and convergent validity (r = .64).

3.5 Testing the feasibility and acceptability of the research components

The feasibility and acceptability of the components of the research were evaluated by piloting the procedures with two patients (in the low secure sample) who were due for discharge. Two patients completed the measures in the order set out in *Section 3.6.* They were asked their opinions of the information sheets, consent forms and measures. Four questions relating to the information and consent forms were: (1) *Is it easy to read* (2) *Is anything not clear* (3) *Does it need any additional information*; and (4) *Do any other changes need to be made*? As the CORE-OM, the ITAQ and Reynolds' Short Form C are all standardised measures, participants were not asked for detailed feedback but, instead, were asked the question: (1) *Were the questions what you had expected from reading the information sheets.* For the Brief CORE-OM Interview participants were asked: (1) *Did it make sense*; and (2) *Are there any other questions that should be asked?* Both participants reported that the consent and information sheets were easy to read, made sense and did not need additional information or changes. Both participants reported that the Brief CORE-OM Interview made sense and neither suggested any additional questions.

A similar procedure was conducted with the primary nurse for both participants. Nurses were asked to comment on nurse information and consent forms (using the same questions that were used for the patients). Nursing staff were then asked three questions focusing on the CORE-OM (Staff Version): (1) *Does it make sense;* (2) *Do the questions seem reasonable for staff to complete;* and (3) *Do any changes need to be made?* Nurse feedback was positive with both stating that the information and consent forms were easy to read, made sense and did not need additional information or changes. With reference to the CORE-OM (Staff Version) one change was suggested. One nurse stated that item 10 of the CORE-OM staff version ("… *has felt like talking to people has felt too much*") should be changed to "*has not felt like talking to anyone*".

However, this was not changed as it is taken directly from the published patient version and consistency between the measures was preferred. No other changes were suggested.

3.6 Study procedure

Initial contact was made with the hospitals through telephone contact with the hospital managers. Meetings were held with clinical directors and hospital managers (High secure = 1, Medium Secure = 2, Low Secure = 3) to obtain permission to conduct the research and to make initial contact with Responsible Clinicians. Meetings with Responsible Clinicians (High Secure = 6, Medium Secure = 4, Low Secure = 4) were conducted to introduce the research and obtain potential participant names (according to the inclusion/exclusion criteria). Meetings were conducted with Ward Managers to discuss the research before then meeting with each potential participant's key nurse to introduce the research s. Potential patient participants were sent an information pack, including an opt-in/out sheet which was given to the primary nurse if the individual wished to take part. This information was obtained through phone calls or visits to key nurses. Patients who did not respond initially were promoted with a brief reminder letter one month after receiving the original and reminded by their primary nurse. Figure 1 (below) shows information on how the sample was obtained while Figure 2 details the numbers who did not meet inclusion criteria.

Figure 1: A CONSORT diagram reporting participant's numbers in each setting.

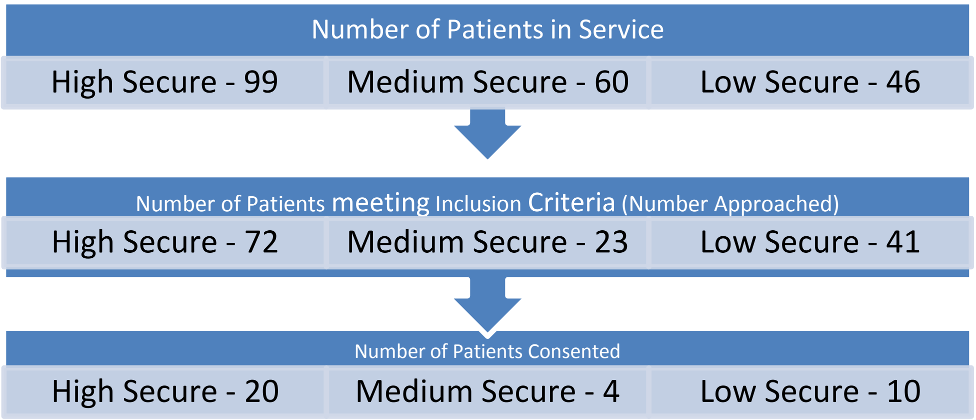
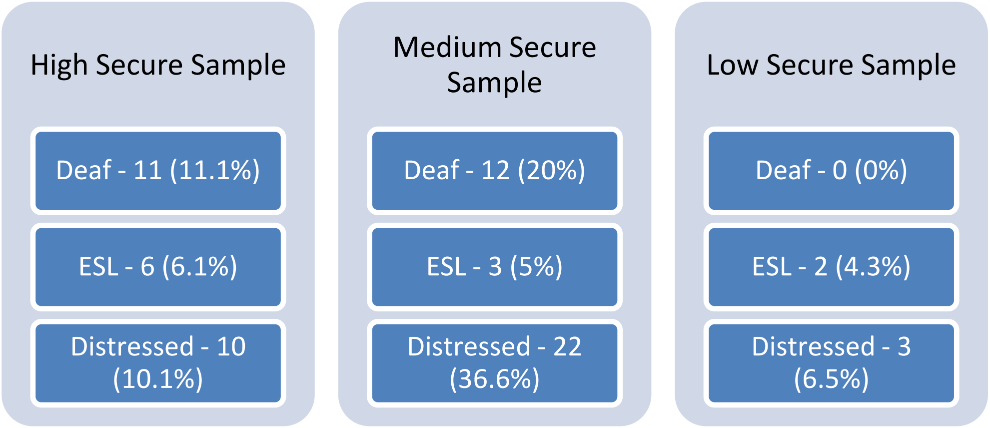


Figure 2: A diagram reporting the numbers who did not meet inclusion criteria (and as a percentage of the total number of patients in the service).



Key: Deaf = Patient is deaf, ESL = English Second Language, Distressed = those who would be too distressed to take part.

After receiving patient opt-in forms the researcher met with the patient to discuss the research and obtain informed consent (allowing a minimum of 24 hours after this meeting before conducting the research according to ethical approval guidance). The primary researcher then made appointments to meet the patient and their key nurse on the same day. Nurses were seen first to obtain any information pertaining to the safety of the researcher (two patients required a staff escort) and to ensure that the key nurse was available for the patient if requested following the research. During the research patients completed the measures face-to-face with the primary researcher (who read all items aloud and documented responses). Administration order is discussed below.

(a) Patient measures: As participants were aware that the CORE-OM was the focus of the study, it seemed important to begin with that measure. In order to obtain the most meaningful feedback, the CORE-OM interview was always conducted immediately after the CORE-OM so that participants had the questionnaire fresh in their minds. A blank copy of the CORE-OM was given to the patient to assist them in responding. All responses were handwritten by the researcher. As the ITAQ has the potential to distress participants (for example, if they are strongly opposed to receiving medication) it felt important to finish with Reynolds' Short Form C which was designed to be pathologically neutral (Andrews & Meyer, 2003) and, as such, less distressing. For these reasons, the measures were completed in the same order for all participants. For patients this was; CORE-OM, followed by the Brief CORE-OM Interview, the ITAQ, and finally the Reynolds' Short Form C of the Marlowe-Crowne Social Desirability Scale (Reynolds' Short Form C).

(b) Staff measures: The staff measures were completed with a participant's key nurse. Again, these were always delivered in the order of CORE-OM (Staff Version), the Brief CORE-OM Interview –Staff Version (so the previous measure was still in mind) and the MIRECC GAF. A blank copy of the CORE-OM was given to the nurse to assist them during the interview. All responses were handwritten by the researcher.

Time taken to complete the CORE-OM (both staff and patients version) was recorded by the researcher. Demographic data (gender, age, diagnosis, current hospital location and length of stay) was obtained from the patient. All participants consented to having the measures feedback to their clinical team. Those who consented to complete the

CORE-OM a second time (to obtain test-retest data) were approached (through their key nurse) within 2 weeks of completing the original version.

3.7 Ethics

Ethical approval was obtained from Nottingham Research Ethics Committee 1 prior to commencing the study (see appendix 1).

3.8 Analyses

3.8.1. Tests of normality

The Shapiro-Wilk Test for normality was used, as recommended for sample sizes under 50 (Shapiro & Wilk, 1965). The results showed many variables were not normally distributed (all being positively skewed). Of the 8 patient-reported variables, only 2 (scores on the CORE-OM *Wellbeing* domain and scores on the Marlow-Crown Social Desirability Scale) were normally distributed. Of the 9 nurse-reported variables, 6 were normally distributed. These were the CORE-OM domains of *Symptoms, Functioning, the Clinical Score,* and Clinical score *minus Risk*; and the MIRECC domains of *Occupational Functioning and Social Functioning.* Outliers were shown for both Patient and Staff reported scores in the *Risk* domain for the CORE-OM. Furthermore, data for these variables were heavily positively skewed with the majority of patients (85.3%) and staff (75.5%) reporting no risk to either self or others. With few patient reported scores being normally distributed and with much of the analysis focusing on patient reported scores non-parametric tests were used throughout. These were Spearman’s Rho (rho) for correlations and Mann Whitney (U) for differences. The CORE-OM Clinical Score was used for comparison with other measures.

3.8.2. Data analysis

(1) Quantitative analysis: Data was analysed using SPSS 16. As this is a study of associations the predominant statistics are correlations (Spearman's Rho). Internal consistency of the CORE-OM was determined by Cronbach's alpha. Acceptability was assessed via the number of non-completed items on the CORE-OM. Differences between mean patient and nurse reported scores on the CORE-OM were analysed using Mann Whitney U

(2) Qualitative Analysis: The semi-structured interviews provided qualitative feedback on the measure. This was analysed using template analysis (King, 1998). Template analysis is a qualitative technique commonly used in healthcare research (i.e. Kent, 2000; King, Thomas & Bell, 2003) and is appropriate for analysing interview data that is not fully transcribed (King, 2004).

The process of analysing and legitimising the data begins with creating a template of high-order themes. King (2004) describes 3 techniques for developing themes: (1) developing *a priori* themes based on relevant research, (2) developing themes after initial exploration of the data, or (3) developing initial codes based on interview questions before refining through exploration of the data. For this analysis *a priori* themes were developed according to the interview questions that are based on the evaluative areas for measures laid out in section 1.6 by Evans, Greenhalgh, and Connelly (2000). Thus, *a priori* themes for both patient and nurse interviews were *Acceptability, Relevance, and Feasibility*. A further shared *a priori* theme was *Changes to treatment* developed from previous research and interview questions. For nurses this was based on the evaluative criteria (Evans et al., 2000) on whether a measure will influence care planning. For patients it was based on the literature on defensive responding (Heilbrun et al., 1990) and relates to the interview question "*Did you worry that answering this questionnaire would affect your treatment or make any difference to how long you stay in hospital"*. A final shared theme was entitled *Understanding.* For nurses it related to the interview question on whether they felt they had a good enough understanding of their patient to answer the questions in the CORE-OM. Whilst high order themes were developed *a priori*, sub-themes were developed after initial exploration of the data and will be described further in the results. Sub- themes are seen to emerge from quotes sorted according to similar topics (Crabtree & Miller, 1999). This leads to the development of the initial template. Coding is then carried out using this initial template for all transcripts, with modifications made wherever necessary. As it is not deemed possible to reach a *final* version of the template (King, 2004) a law of diminishing returns is applied when continuing to recode data that does not lead to further enrichment of data. Finally, Waring and Wainwright (2008) recommend calculating the frequency of quotes relating to each narrow theme. During the interpretation, quotes from participants should be used to describe the themes (King, 2004).

3.9 Determining sample size

The required sample size of patients was determined through the use of GPower (Faul & Erdfleder, 1992). To date, there appears to be only one paper making direct comparisons between patient and key worker reports on the same measure, reporting an agreement level of .4 (Hunter et al., 2004). Thus, if this value is retained on the basis of methodology, the power analysis yields a required sample of 34 (r = .4, p = .05, power = .80). However, as this study intended to use intraclass correlations, a second power analysis was conducted on the basis of statistical analysis informed by the correlations between clinician and patient for anxiety (Morlan & Tan, 1998) and depression (Preston

& Harrison, 2003). This yielded a sample size of 25 (r = .46, p = .05, power = .80). Thus, a sample size between 25 and 34 was required. As forensic inpatients are difficult to recruit (Fitzpatrick et al., 2010) a large sample was approached to ensure the minimum sample size was obtained.

4. Results

Results (reported at the 2-tail level of significance) are discussed according to the aims described in *Section 2.8.* For clarity Aim 4 will be discussed in *Section 4.5.*

Part 1: Quantitative results

4.1. Aim 1: Psychometric data

i. *Reliability:* Internal consistency for the CORE-OM (PV) was α= .81. Test-retest correlations, taken from a sample who completed the CORE-OM twice (n = 11), was rho = .66. The internal reliability for the CORE-OM (SV) was α = .87.
ii. *Acceptability:* Mean completion time for the CORE-OM (PV) was 5.56 minutes (SD = 1.86), and 5.68 minutes (SD = 2.25) for the CORE-OM (SV). From a total item pool of 1156 patient-completed items, only a single item (0.08%) was missing. For staff a total of 7 (0.60%) out of 1156 items were missing.
iii. *Mean domain and clinical CORE-OM scores:* Mean scores reported by patients and nurses (Table 1) are in the low level (6-9) range. To provide comparisons Table 2 shows clinical scores for non-clinical, clinical and a prisoner population (McCloskey, 2001) alongside the patient reported scores from this study (secure hospitals).

Table 1: Mean domain and clinical scores on the CORE-OM PV and SV (for patients and nurses)

Domain	**Perspective** Patient n = 34 Nurse n = 34	**Mean**	**Std. Deviation**	**Minimum**	**Maximum**
Wellbeing	Patient	7.99	5.95	0.00	22.50
	Nurse	8.19	4.60	0.00	30.00
Symptoms	Patient	9.58	6.04	0.00	22.50
	Nurse	8.69	5.20	0.00	20.00
Functioning	Patient	8.06	5.66	0.83	20.00
	Nurse	7.98	5.51	0.00	20.83
Risk	Patient	0.49	1.39	0.00	6.67
	Nurse	0.78	1.75	0.00	8.33
Clinical Score	Patient	7.16	3.91	1.18	15.00
	Nurse	6.94	3.95	0.00	16.18
Clinical Score (- Risk)	Patient	8.59	4.60	1.43	18.21
	Nurse	8.27	4.81	0.00	19.29

Table 2: Means and Std Deviations non-clinical, clinical, prisoner, and secure hospital samples.

CORE-OM (PV)	Non- Clinical Mean (S.D)	Clinical Mean (S.D)	Prisoner Mean (S.D)	Secure Hospitals Mean (S.D)
Wellbeing	9.1 (8.3)	23.7 (9.6)	16.3 (8.8)	7.99 (5.95)
Symptoms	9.0 (2.0)	23.1 (8.8)	15.3 (9.1)	9.58 (6.04)
Functioning	8.5 (6.5)	18.6 (8.4)	13.4 (7.7)	8.06 (5.66)
Risk	2.0 (4.5)	6.3 (7.5)	3.9 (6.5)	0.49 (1.39)
Clinical Score	8.8 (6.6)	21.2 (8.1)	16.2 (8.3)	7.16 (3.91)
Clinical Score minus risk	7.6 (5.9)	18.6 (7.5)	14.1 (7.7)	8.59 (4.60)

4.2. Aim 2: Determining correlations between patient and nurse responses on domains of the CORE-OM

Patient clinical scores in domains of the CORE-OM (PV) were correlated with the respective domain on the nurse-completed CORE-OM (SV). No significant correlations between patient- and nurse-reported CORE-OM scores were found for *Wellbeing* (rho = .013, n = 34, p = .94), *Symptoms* (rho = .26, n = 34, p = .13), *Risk* (rho = .063, n = 34, p = .72), *or CORE-OM clinical score minus risk* (rho = .30, n = 34, p = .09). The correlation between patient- and nurse-reported scores on the *Functioning* domain were significant (rho = .38, n = 34, p = .04) and the correlations for *Clinical Score* approached significance (r = .33, n = 34, p = .06). Scatter-plots, with linear regression lines, for all domains are presented in Figures 3–8 to indicate the direction and strength of the correlations.

Figure 3: Correlations between patient and nurse reported scores on the Wellbeing domain

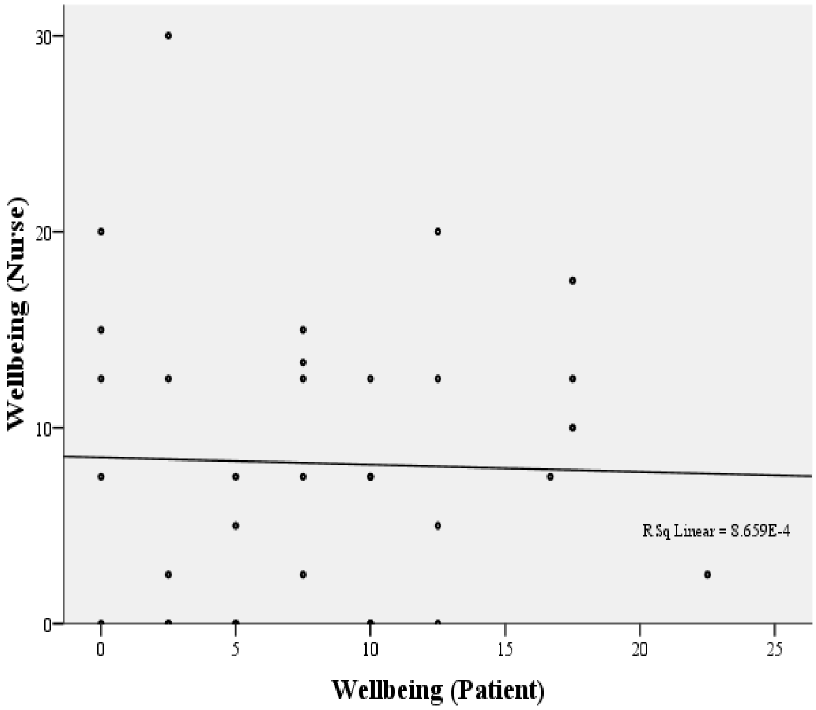


Figure 4: Correlations between patient and nurse reported scores on the Symptoms domain

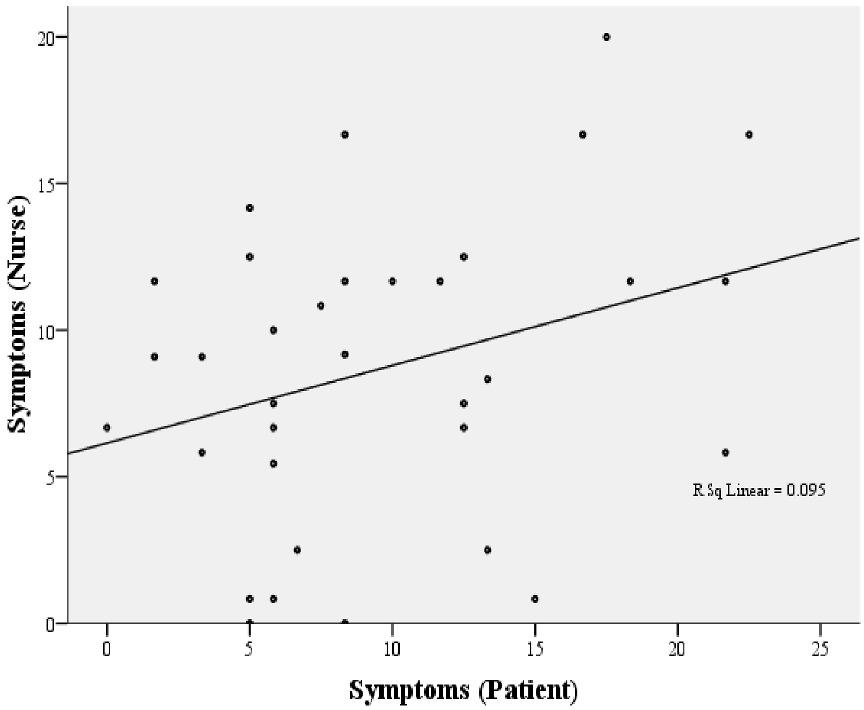


Figure 5: Correlations between patient and nurse reported scores on the Functioning domain

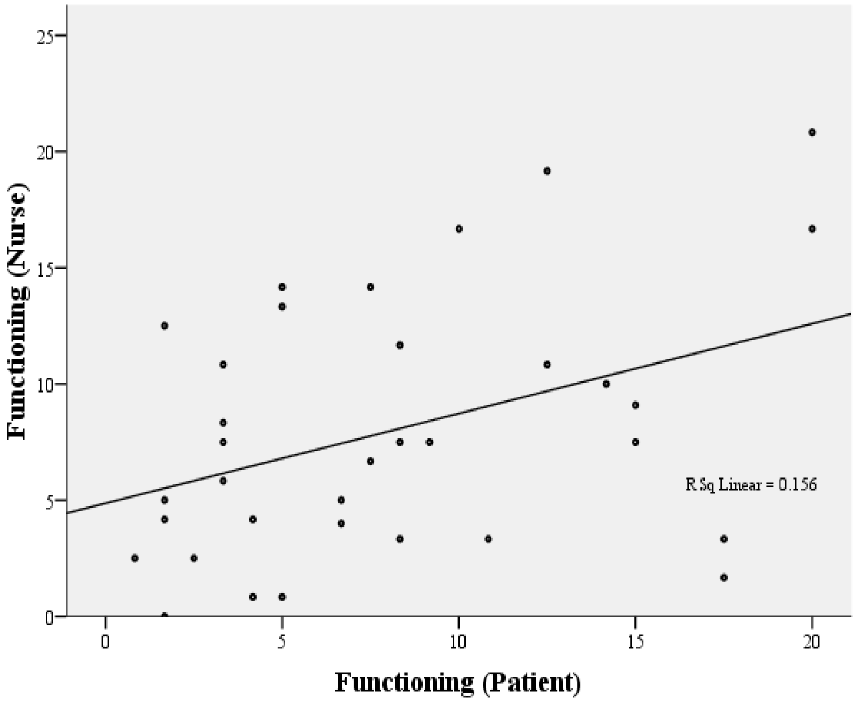


Figure 6: Correlations between patient and nurse reported scores on the Risk domain

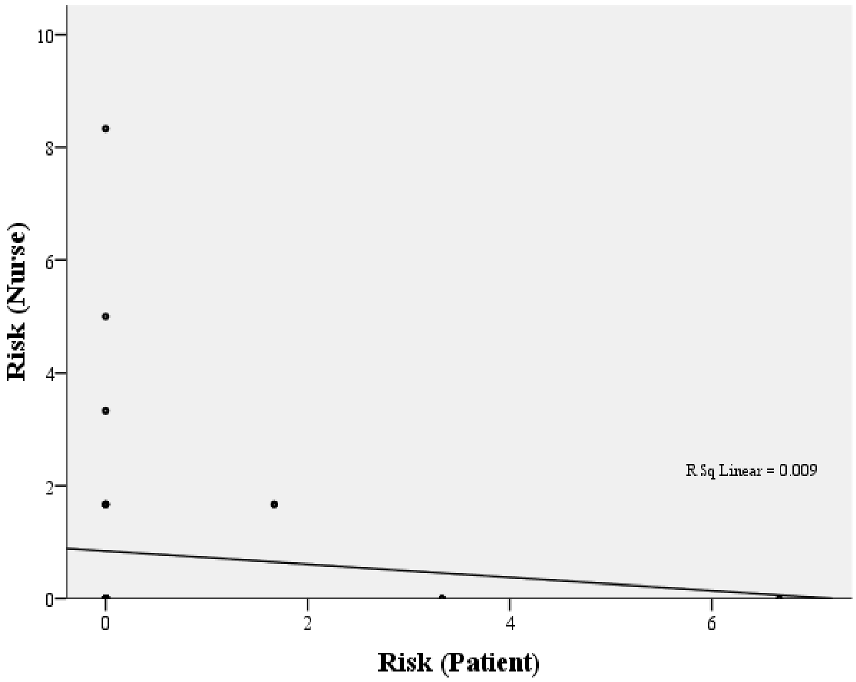


Figure 7: Correlations between patient and staff clinical scores (CORE-OM)

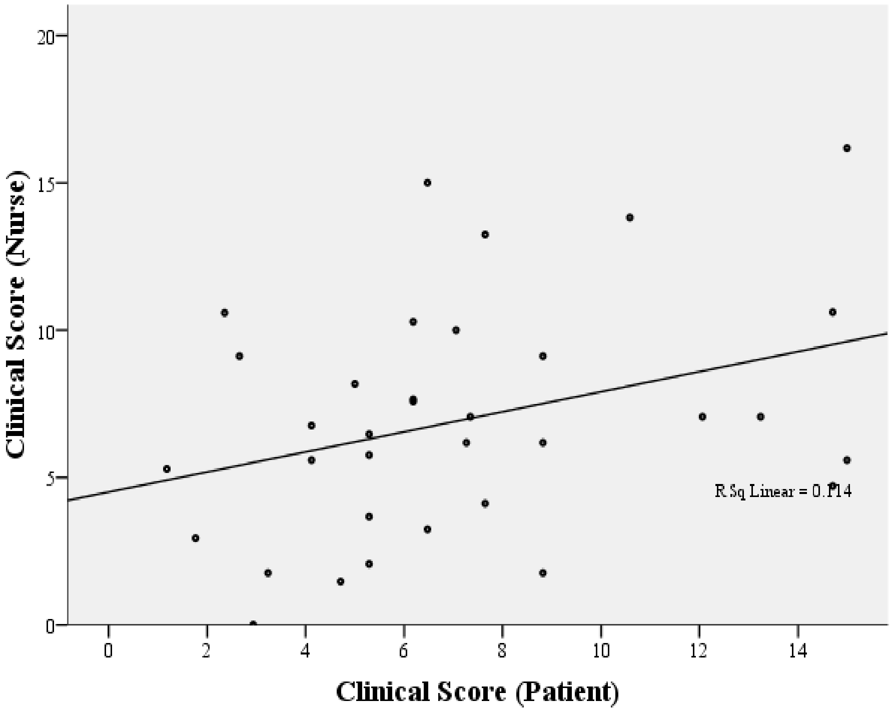
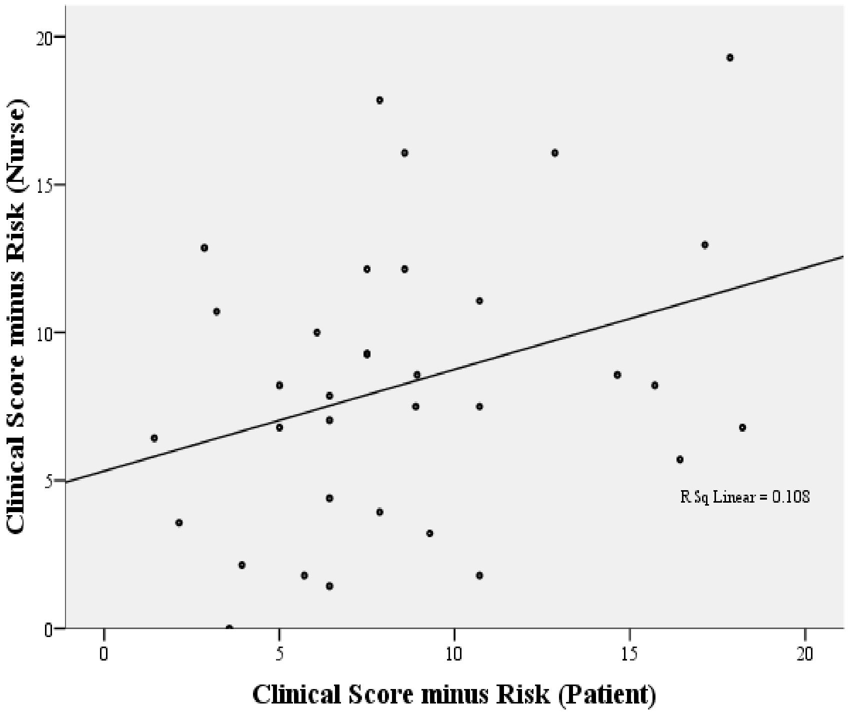


Figure 8: Correlations between patient and nurse reported clinical scores minus risk (CORE-OM)



4.3. Aim 3: Assessing the relationship between patient level of functioning (nurse-rated) and severity of difficulties as rated on the CORE-OM

The data were analysed to determine whether levels of *Occupational, Social and Symptomatic functioning* (nurse-rated) were significantly correlated with domain scores on the CORE-OM (Patient and Staff Versions). The correlations are shown in Table 3. *Occupational Functioning* was not significantly correlated with any patient- or staff-reported domain scores. *Social Functioning* was significantly correlated with the *Functioning* domain as rated by staff (rho = -.48, n = 34, p = .004). *Symptomatic Functioning* was significantly correlated with the *Functioning* domain for staff (rho = -.40, n = 34, p = .02) and patients (rho = -.37, n = 34, p = .03). *Symptomatic Functioning* was also significantly correlated with the *Clinical score* for Staff (rho = -.38, n = 34, p = .02) and Patient (rho = -.41, n = 34, p = .01); and the *Clinical score (minus Risk)* for Staff (r = -.36, n = 34, p = .03) and Patients (rho = -.40, n = 34, p = .02). Due to space constraints scatter-plots are shown only for significant correlations (see Figures 9 to 15).

Table 3: Correlations between nurse-rated functioning (MIRECC GAF) and CORE-OM scores

	Nurse-rated Functioning (MIRECC GAF)		
	Occupational	Social	Symptomatic
CORE-OM Score (SV)			
Wellbeing	.02	-.04	-.11
Symptoms	.11	-.04	-.29
Functioning	-.17	-.48 **	-.40*
Risk	-.01	-.10	-.12
Clinical Score	-.06	-.29	-.38*
Clinical Score -Risk	-.05	-.28	-.36*
CORE-OM Score (PV)			
Wellbeing	-.06	.07	-.29
Symptoms	.13	.14	-.25
Functioning	-.32	-/08	-.37 *
Risk	.07	-.02	-. 27
Clinical Score	-.09	.04	-.41 *
Clinical Score -Risk	-.11	.05	-.40*

* correlation significant at the .05 level (2-tailed), ** correlation significant at the 0.01 level (2-tailed)

Figure 9: Correlations between nurse rated Social functioning (MIRECC GAF) and nurse rated Functioning (CORE-OM)

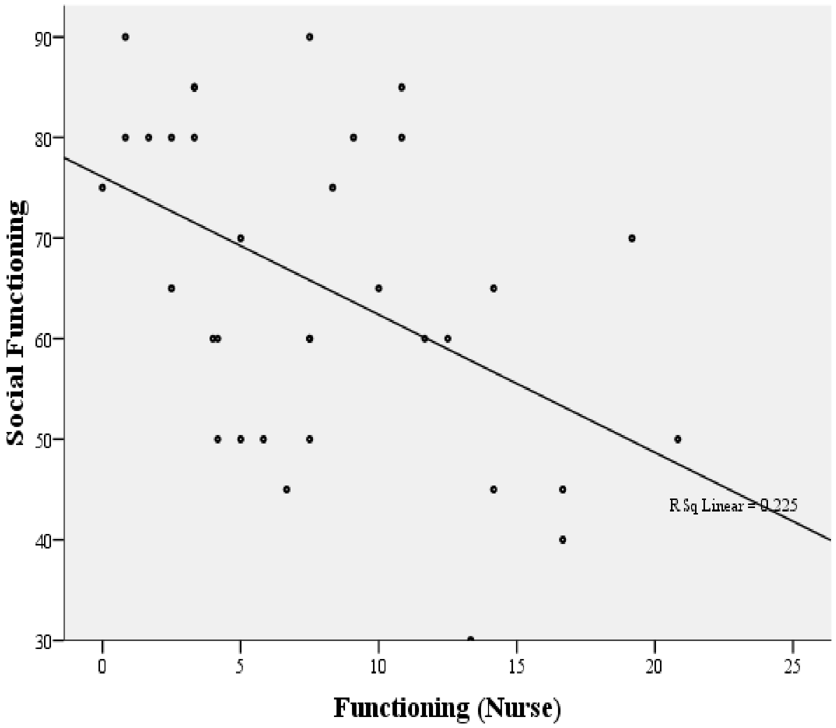


Figure 10: Correlations between nurse rated Symptomatic functioning (MIRECC GAF) and nurse rated Functioning (CORE-OM)

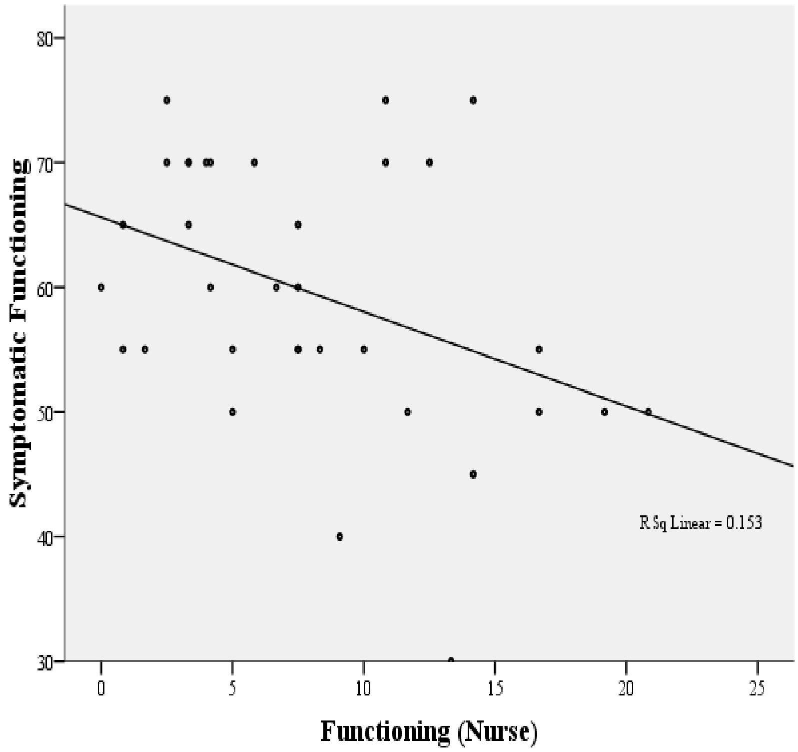


Figure 11: Correlations between nurse rated Symptomatic functioning (MIRECC GAF) and patient rated Functioning (CORE-OM)

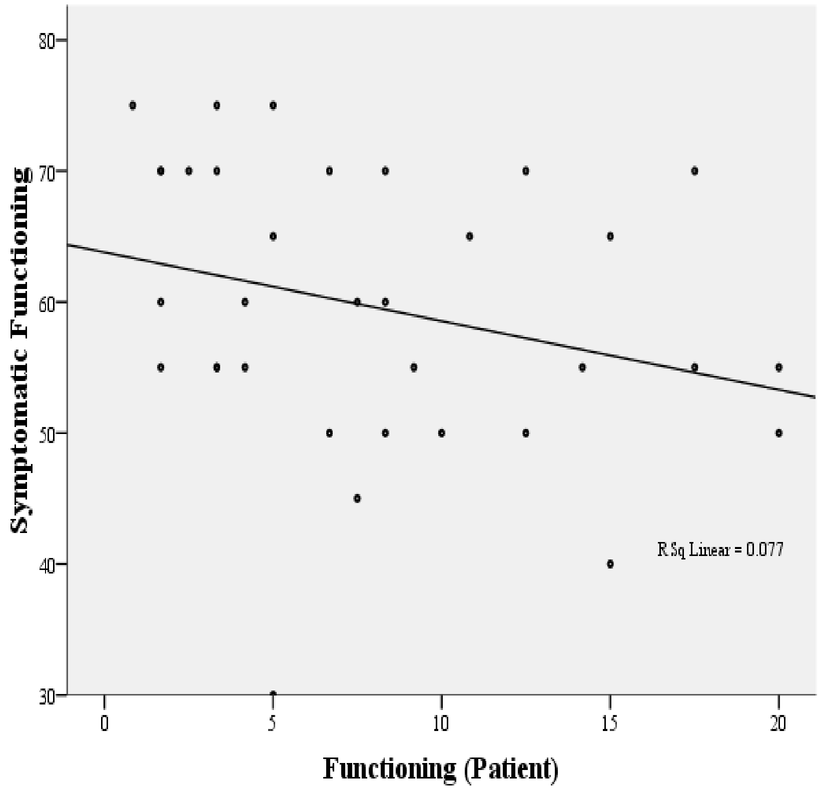


Figure 12: Correlations between nurse rated Symptomatic Functioning (MIRECC GAF) and nurse rated Clinical score (CORE-OM)

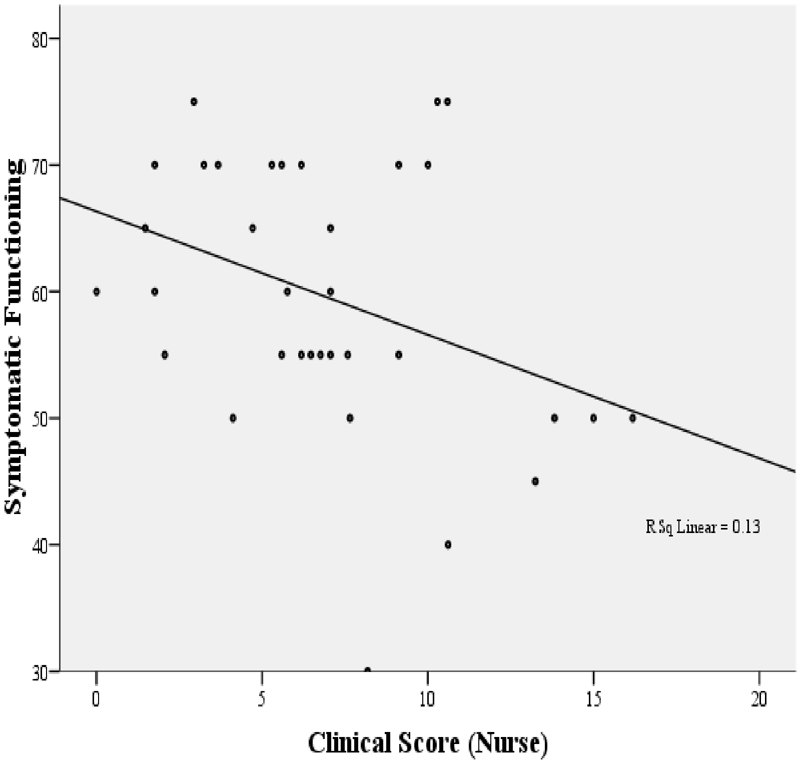


Figure 13: Correlations between nurse rated Symptomatic Functioning (MIRECC GAF) and patient rated Clinical score (CORE-OM)

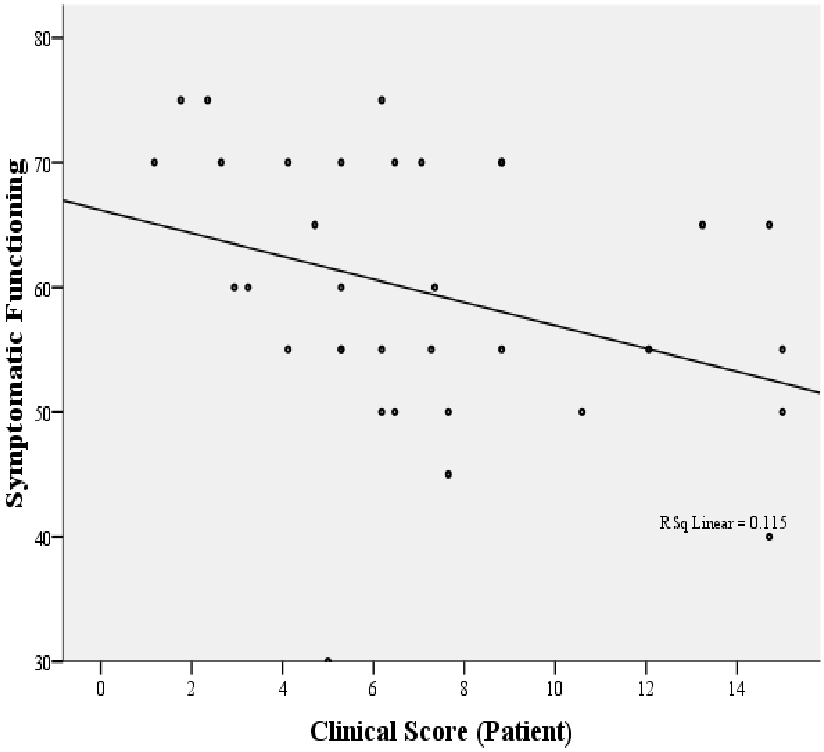


Figure 14: Correlations between nurse rated Symptomatic Functioning (MIRECC GAF) and nurse rated Clinical score minus risk (CORE-OM)

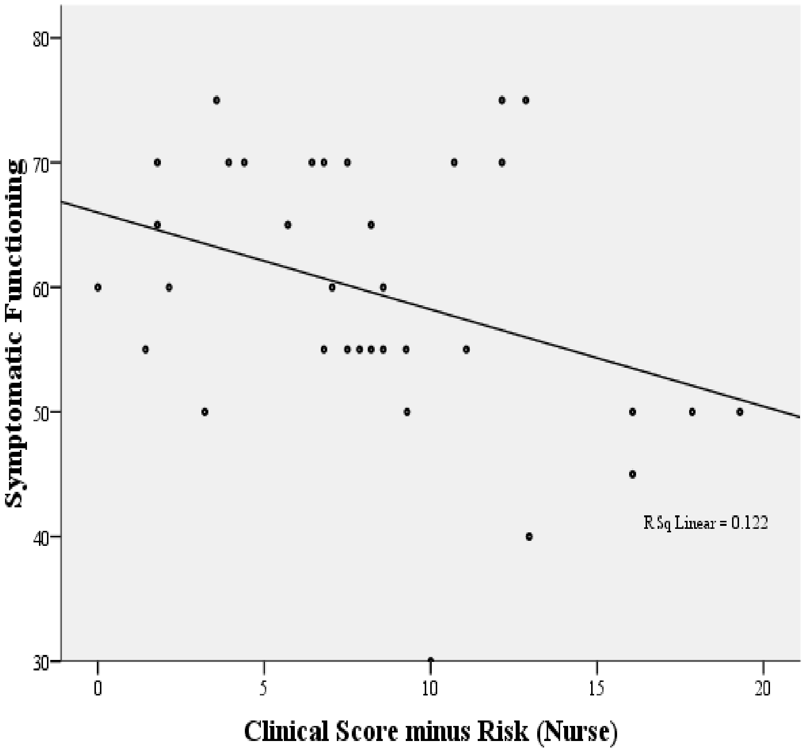
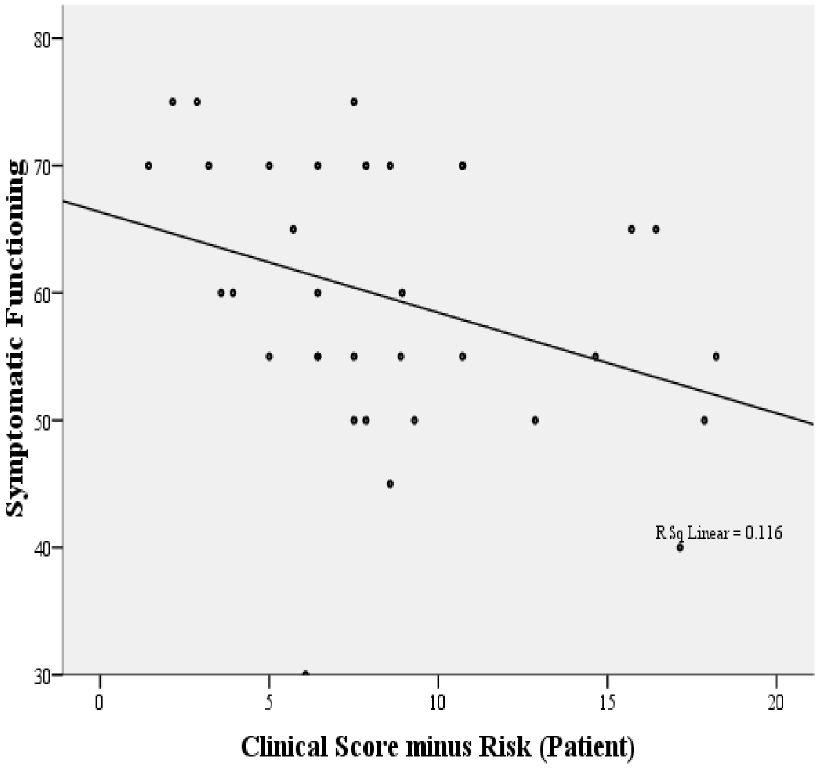


Figure 15: Correlations between nurse rated Symptomatic Functioning (MIRECC GAF) and patient rated Clinical score minus risk (CORE-OM)



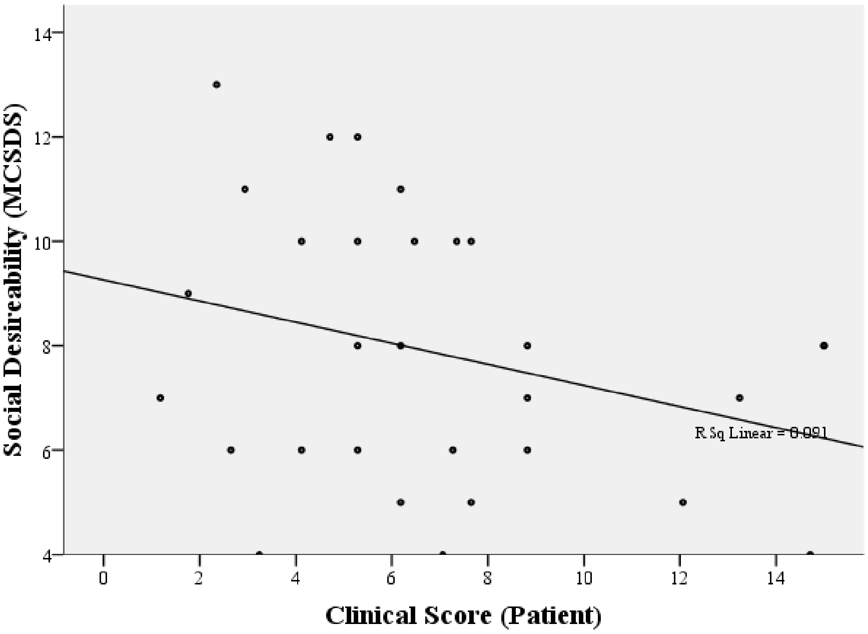
4.4. Aims 5, 6 and 7: Investigating response styles

In light of the small number of significant correlations obtained between patient- and nurse-reported scores on the domains of the CORE-OM, further analyses to investigate the impact of response style were conducted.

i. Aim 5: Assessing differences between nurse- and patient-reported CORE-OM scores (Malingering). No significant differences were obtained between patient- and staff-reported scores for *Wellbeing* (U = 570.50, p = .93), *Symptoms* (U = .555.50, p = .78), *Functioning* (U =575.00, p = .97), *Risk* (U = 514.50, p = .27), *Clinical score* (U =577.00, p = .99) or *Clinical score -minus Risk* (U = 572.000, p = .94). Visual inspection of the mean domains and clinical scores reported in Table 1 shows that patient scores are slightly lower than nurse reported scores for *Wellbeing* and *Risk;* and slightly higher for *Symptoms, Functioning, Clinical Score* and *Clinical Score minus Risk.*

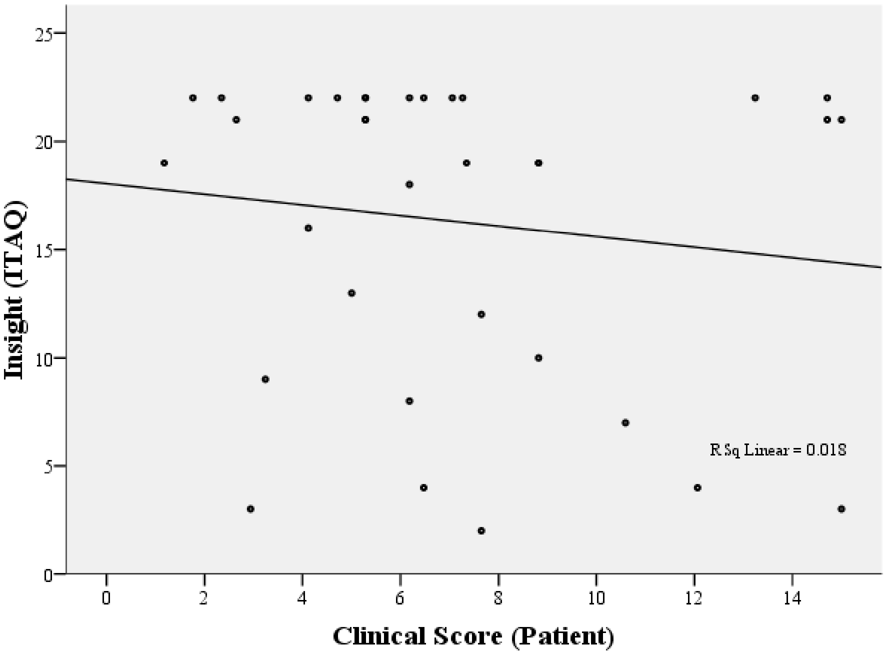
ii. Aim 6: Assessing the relationship between Clinical CORE-OM (PV) scores and social desirability (Defensiveness). No significant correlations were obtained between scores on the MCSDS (for Social Desirability) and Total CORE-OM scores for patients (r = -.28, n = 34, p =.13). A scatter-plot is shown in Figure 16.

Figure 16: Correlations between patient reported Clinical score and social desirability (MCSDS)



iii. <u>Aim 7: Assessing the relationship between Clinical CORE-OM (PV) scores and Insight.</u> No significant correlations obtained between scores on the ITAQ (for insight) and mean item CORE-OM scores for patients (r = -.19, n = 34, p = .28). A scatter plot is shown in Figure 17.

Figure 17: Correlations between patient reported Clinical score and insight (ITAQ)



Part 2: Qualitative results

4.5. Acceptability and feasibility of using the CORE-OM in secure settings

Data from the Brief CORE Interviews (staff and patient) were analysed using Template Analysis (King, 1998). The themes identified prior to analysis (in line with the questions) were *Acceptability, Relevance, Feasibility, Treatment,* and *Understanding* (see Section 3.8.2 for details). After using this template to code responses, one additional shared theme emerged and was titled *Suitability*. From here several sub-themes and levels within these sub-themes emerged, derived through a process of immersion and crystallization. Once developed the template was audited by a member of the research team (M.B). The results are discussed in terms of high order themes, sub-themes, and levels (King, 1998). The themes for patients and nurses are introduced in Figures 18 and 19. Sub-themes can be seen through the text in bold, with levels within the subthemes shown in bold italic. Due to space constraints limited quotes are shown in the text. However, the completed templates (with all responses) are presented in Appendix 3. Patient (n = 34) and Staff (n = 24) responses are discussed under the same themes. There are fewer staff responses as 3 staff members were key nurses for several patients.

Figure 18: Higher order themes and sub-themes for patients

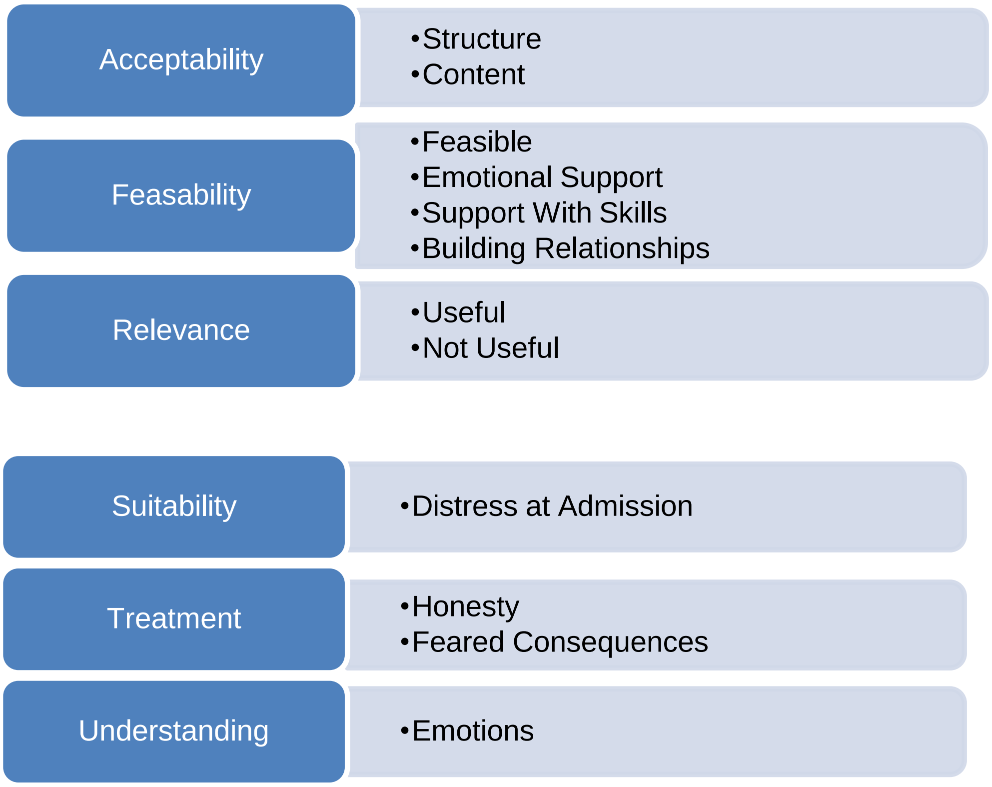
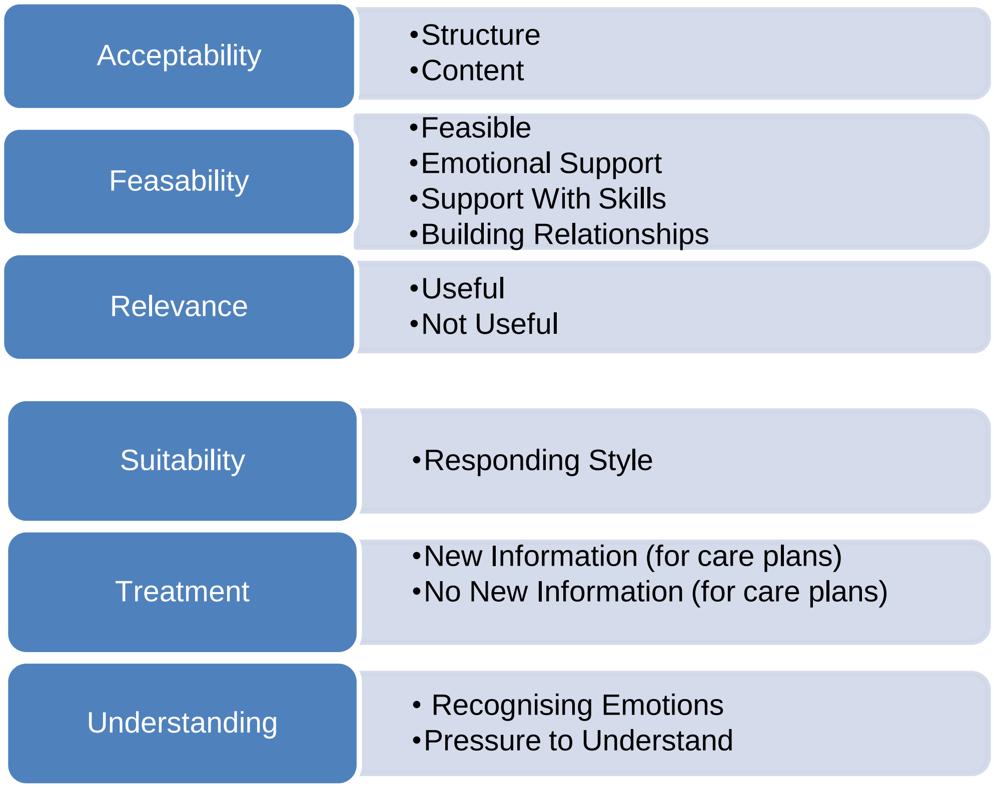


Figure 19: Higher order themes and subthemes for nurses



Higher order theme: Acceptability

Two sub-themes that emerged for both patients and nurses were **Structure** and **Content**. For patients the **structure** mostly (n = 9) related to ***simplicity*** with comments such as “the questions were easy for me to understand” although 4 patients commented about ***difficulties with categorising*** (“sometimes it was hard to decide between ‘Sometimes’ and ‘Often’”). In relation to **content** the majority of patients reported the questions made them feel ***comfortable*** (n = 9) with comments such as “... this seemed less intrusive than previous questionnaires”. Others (n = 6) commented on finding the questions ***interesting or useful*** (“it’s interesting, it gives an insight into an individual”). However, some (n = 3) reported difficulties relating to the ***potential emotional impact*** (“... it brought out a bit of grief”) or ***lack of positivity*** (n = 3).

Nurses also reported themes around ***simplicity*** (n = 6; “it’s very easy to answer”), although more (n = 8) reported ***difficulties with categorising*** (“it can be hard to be so precise in the timings”). With the **content** theme nurses commented on the ***variety*** of topics covered by the CORE-OM (n = 5; “it looks at everything from functioning to behaviour”), as well as the ability to use the scores to provide ***comparisons*** (n = 3; “you could use it to detect changes”). However, 3 staff members also commented on a ***lack of positivity*** (“sometimes it feels like it pathologises things we all feel at times”). An additional sub-theme theme transpired for nurses (n = 3) around the **accessibility** of the CORE-OM to all profession, with comments such as “a lot of these things are done by specialist services who ...rarely see the patient”.

Higher order theme: Feasibility

Within this theme, the majority of patients (n = 26) and nurses (n = 20) felt that the patient should complete the CORE-OM with a staff member. Themes for both patients and nurses emerged around needing **emotional support** (patient n = 10, nurse n = 8), **support with skills** (patient n = 13, nurse n = 8) and **building relationships** (patient n = 3, nurse n = 4). Comments from patients included "it's better with a key worker to talk about things" (emotional support), "...my reading and writing aren't very good" (skills) and "...it would help build rapport" (building relationships). For nurses comments included "he'd get too upset to do it alone" (emotional support), "his literacy skills aren't great" (skills) and "it would be useful as the first line to starting the therapeutic relationship" (building relationships).

Higher order theme: Relevance

Sub themes for both patients and nurses related to the CORE-OM being either **useful** or **not useful**. For patients the CORE-OM was deemed to be **useful** as a way of ***sharing views*** (n = 9; "it would be helpful when you want to put across your insight"), and ***sharing emotions*** (n = 8; "it would help me talk to my nurse"). Responses coded as **not useful** related to a ***preference for talking*** (n = 6; "I'd rather someone talked to me than did paperwork"), ***individual differences*** (n = 3; "people are too different") and a perceived ***lack of need*** (n = 6; "I can cope without the help of the hospital").

More nurse (n = 13) than patient responses could be coded into the **useful** theme with comments relating to helping the patient to ***share emotions*** ("it would help a patient express their feelings"), and ***helping nurses to structure patient sessions*** ("it would focus onto the way the patient is feeling"). For responses coded into the **not useful** theme comments related to a ***preference for talking*** (n =5; "sometimes ... you are overburdened with tick boxes”), ***lack of focus on offending*** (n = 2; “it didn’t cover the things I expected it to for those who have a history of violence”) and ***psychosis*** (n = 3; “it doesn’t focus on severe mental health problems”).

Higher order theme: Suitability

Due to a difference in the interview questions, the sub themes for patients and nurses were different in this section. Although 10 patients reported that it would be appropriate for patients to complete the CORE-OM at admission, most comments could be coded into a theme related to ***distress at admission*** (n = 24). Comments here included “I couldn’t have done it ... my head was gone”, “early on my problems were too big, it wouldn’t have been relevant” and “I wouldn’t have been honest, I’d have been too frightened”. While 14 nurses stated that the CORE-OM was **suitable,** others expressed concern about patient **responding style** such as ***malingering*** (n = 2; “some might do it just to get attention”), ***defensiveness*** (n = 5; “they would probably lie or mask their answers”) and ***insight*** (n = 3; “some have little self awareness and may not be able to answer it”).

Higher order theme: Treatment

Again, sub themes were different for patients and nurses. All patients (n = 34) reported that they felt able to be **honest** and open in their responses. However, a theme emerged for 11 patients around **feared consequences** of revealing difficulties. Comments here included “I’ve learnt that, in this system, if you reveal things like depression you are punished for it. My room was stripped, I was placed on observations. I’ve learnt never to reveal those things in a place like this...”, “I’d want to tell the truth but they’d just put me on obs(ervations)” and “other patients tell me ‘don’t tell them!’ so you can worry”.

Most nurse responses could be coded into a theme that the CORE-OM would provide **new information** and aid care planning by ***raising new issues*** (n = 5; “it might help pull out things you weren’t aware of”), ***providing structure*** (n = 3; “it might help give a rationale for things”), ***aiding clinical team work*** (n = 4; “it might help generate discussion”) and ***helping to get to know someone new*** (n = 3; “it is probably more useful with someone you don’t know”). Nine responses were coded into the theme of **no new information** and included comments such as “I know how patients are without the structure of a questionnaire” and “my patient is good at explaining his feelings”.

Higher order theme: Understanding

All patients reported that they had a good enough understanding of their emotions to answer the questions (n = 34), although two patients commented “yes, just about” and “with some of them (questions) but some not”. While the majority (n = 15) of nurses reported that did know their patients well enough to answer the questions in the CORE-OM, an additional theme emerged around **recognising Emotion** (n = 9) with comments including “it’s hard to know what someone is feeling” and “I can only comment on the behaviour shown”. Finally, most (n = 18) nurses did not report feeling under **pressure to understand** their patients (i.e. “it’s just my job” and “I want to get it right”). However, some (n = 6) reported some pressure with comments such as “it can affect his treatment so I want ... to get it right” and “you feel like you have to answer, even if the question doesn’t seem right”.

4.6 Summary of qualitative findings

Five of the 6 higher order themes derived from the original questionnaire were maintained for both patients and nurses. One additional higher order theme of (suitability) emerged. The data will be discussed further in the following section.

5. Discussion

5.1 Structure

Discussion of the results will occur in 2 sections. The first section will focus on the results relating to using the CORE-OM in secure hospitals. This will be achieved by discussing the reported mean domain and clinical scores, the correlations between the patient and nurse CORE-OM scores, and the results relating to patient responding style. In the second section the discussion will be structured according to the assessment criteria laid out by Evans et al. (2000; Section 1.6) as a way of determining whether the CORE-OM (Patient-Version) is a suitable measure for use in secure hospitals. These criteria are: psychometric properties, stakeholder perspectives, feasibility, and utility. Finally, clinical implications, limitations, and suggestions for further research will be discussed.

5.2. Using the CORE-OM in secure hospitals

5.2.1 Means

Notably, the scores in the secure hospital sample are lower than all others, including the non-clinical sample. While it has been suggested that low scoring is not uncommon in complex settings (Barkham et al., 2005), these means are notably lower than other populations and are deemed to be in the ‘low level’ range. There appear to be several potential explanations for this. Firstly, studies investigating mental health difficulties in complex settings suggest that psychosis is a more frequent diagnosis than anxiety, depression or trauma (Murray, 1996; Smith et al., 1991; Taylor, 1998). It may be that overall scores on the CORE-OM are low as they do not directly measure psychotic experiences. However, scores within the Risk domain were also low. Considering admission to high secure hospitals rests on the criteria that the patient is a “grave or immediate risk” (Rampton Admission Guidelines), the low risk scores in this study are questionable and may occur as a result of the limited time period for scoring (1 week). It is also possible that the CORE-OM failure to detect risk is a reflection of risk being controlled within the hospital setting. Thus, it is possible that the low scores are an accurate reflection of the patient’s difficulties within that time period.

Secondly, it should be noted that some patients reported a fear of the consequences when reporting difficulties in secure hospital settings. Although no significant correlations are shown between the measures of social desirability (MCSDS) and clinical CORE-OM scores it is possible that this measure was not able to accurate detect defensive responding. Thirdly, it should be acknowledged that the response rate for this research was relatively low and the consenting sample may not reflect the secure hospital populations. This is potentially supported by the CORE-OM scores provided by nurses.

5.2.2. Comparing staff and patient reported measures.

The only significant correlation between patient and staff ratings on the CORE-OM pertained to the functioning domain (rho =.38), although the clinical score was approaching significance (rho = .33). This appears to be in line with previous research that has shown staff are more likely to identify tangible needs in their patients (Najim & McCrone, 2005; Slade 1994). This suggestion is further supported by the staff-reported theme on identifying patient emotions that suggested nursing staff feel better able to report observable difficulties.

There are several possible explanations for the lack of correlations between staff and patient ratings. Forensic mental health nurses report that limited resources and changes in shift patterns negatively impact on their ability to be fully involved with their clients (Dale & Storey, 2004). These difficulties were apparent during the research when many participants were seen in the evening as this was the only time both key worker and patient were together. With research suggesting that nurses feel less involved with their patients when working night shifts (Dale & Storey, 2004), it is possible that this had an impact on their ability to assess client difficulties on the CORE-OM.

The literature suggests two main styles of forensic nursing care: paternalistic and relational (Gildberg, Elverdam & Hounsguaard., 2010) with paternalistic care led by observations and rule enforcement. It has been suggested that nurse interactions with forensic patients focus on explaining consequences of behaviour and communicating about functioning in daily life (Rask & Lavander, 2000) while nursing entries in patient clinical notes document observations of patient behaviour with little detail of patient needs (Parker & Gardener, 1992). Taking this research into consideration it is possible that the significant correlations for functioning are a reflection of a paternalistic nursing style as staff aim to observe difficulties rather than discover them through discussion with the patient. However, it should be acknowledge that the sample size in this study was small and it is possible that stronger correlations would have been obtained with a larger sample.

5.2.3 Patient responding style

While the results do not appear to show signs of irrelevant responding (evidenced by the internal reliability scores) or malingering (with no significant differences between patient and nurse-reported scores), the clinical scores are lower than one might expect. While the results showed no significant correlation between social desirability and clinical CORE-OM, a negative correlation of .28 suggests that CORE-OM scores were slightly lower for those with higher levels of social desirability. Again, this weak correlation may be a result of the small sample size used. Following on from the patient sub-theme of feared consequences it is possible that some patients hide their difficulties from their key workers leading to the low scores reported by both patients and nurses.

5.3. Assessing the suitability of using the CORE-OM (PV).

5.3.1. Psychometric properties

The results showed the CORE-OM (PV) to have good internal consistency and acceptable test-retest reliability. Although it was beyond the scope of this study to fully determine the validity of the CORE-OM, the significant correlations between patients reported scores for the CORE-OM domain of functioning and staff assessment of symptomatic functioning on the MIRECC GAF are indicative of convergent validity. However, the relatively small sample size means findings relating to validity need to be interpreted with caution. Overall, the psychometric data obtained suggests that the CORE-OM is reliable when used with forensic patients but further work is needed to determine validity.

5.3.2. Stakeholder perspective

The data taken from the Brief CORE-OM interviews suggest that the CORE-OM captures the views of the patients although some difficulties were reported. The findings suggest that the CORE-OM does capture stakeholder perspectives in secure hospitals, although there is the need to support patients in categorising their responses.

5.3.3. Feasibility

Patients in this study were able to complete the CORE-OM with the researcher in less than 6 minutes. While this does not indicate the feasibility of the patient completing the measure alone, the majority of patient and staff agreed that the patient should complete the CORE-OM with a key worker. Patients reported that it would not be useful to complete the CORE-OM at the point of admission. This is consistent with previous studies showing very low correlations between staff and patient measures at the point of admission (Lindenmayer, Kay & Plutchik., 1992). The clinical implication is that standardised measures may not be useful for recently admitted patients. Although the CORE-OM was scored by the researcher, this is also a short process and, if required, a computerised scoring package can be obtained to assist staff. Training is not needed to administer or score the CORE-OM and it is in the public domain and free to use.

5.3.4. Utility

Patient opinion on the utility of the questionnaire was divided. A total of 17 reported that it would be useful for sharing emotions and views while 15 explicitly stated that it would not be useful due to lack of perceived need, individual differences and a preference for talking. While the majority of nurses reported that the CORE-OM would be useful, they also expressed concerns relating to a preference for talking. Thus it is clear that while the CORE-OM may be a useful measure in forensic services, it should not be offered as a replacement for discussion with the patient. The qualitative feedback suggested that staff were keen to use measures that address risk and severe mental health difficulties in more detail. Although the CORE-OM is designed to detect risk, staff concern about its use appears to be valid as the majority of staff and patients reported that the patient was at not a risk to themselves or others. Thus, there appears to be the need to use the CORE-OM alongside a formal forensic risk assessment.

5.4. Clinical implications

While the results suggest that the CORE-OM is acceptable, feasible and captures the views of stakeholders; and that forensic inpatients self reports of difficulties with wellbeing, psychological symptoms, and functioning are not unduly influenced by a malingering or irrelevant responding style, the low clinical scores appear at odd with the complex population sampled. If the CORE-OM is to be used in secure hospitals it should be used alongside measures of risk and psychosis, should not be used at the point of admission, and should be conducted alongside a key worker.

Perhaps the most interesting clinical implication relates the lack of significant correlations between patient and staff rated measures. It is possible that nursing care in forensic services is restricted either by practical difficulties with resources and shift patterns, or by a focus on paternalistic above relational care. Thus, there is the need to ensure that forensic patients have regular contact with their key nurse, but also that forensic nurses are trained to recognise emotional needs.

The suggestion that some forensic patients fear the consequences of reporting their difficulties is of great interest clinically and corresponds to the idea of paternalistic care in forensic settings. This suggests the need for staff to develop alternatives to managing patient distress through observations. It is likely that a shift to relational care could address this.

5.5 Limitations

There are several limitations to this study. Although the sample accessed was large, the response rate was low at 28% (high secure), 17 %(medium secure) and 24% (low secure). As previously noted, accessing participants in forensic settings in extremely complex and there is an urgent need to investigate the use of self-report measures in forensic samples (Fitzpatrick, 2010). The present study was designed as a feasibility study and provides an estimate of the expected response rate that can inform the design of future studies. A large number of correlations were conducted on a small sample. While this may lead to the possibility of Type I errors, this was tolerated as recommended for research assessing associations between patient and clinician rated measures (Huppert, Smith & Apfeldorf., 2002). Although the sample is large enough to detect moderate to large correlations, it may not have been large enough to detect small correlations. Post hoc power analysis, using the correlation reported between patient and nurse ratings of functioning on the CORE-OM, shows the power to detect significant correlations was .70 rather than the preferred .80 ($r = .37$, $p = .04$, power = .70). The sample obtained for test-retest reliability was smaller still. The reduction here was primarily due to the restricted time period for completing the CORE-OM a second time. Imposing a follow up of 2 weeks meant that many patients were not available for follow-up. Thus, although the information presented for test-retest reliability gives an indication as to the reliability of the CORE-OM in forensic services, a larger sample would be required to confirm the accuracy of the results.

Patients were explicitly informed that their responses would be accessible to their clinical care team (with 16% of those who refused to take part declining for this reason) and this may have impacted on the response rate and the results. Considering the qualitative findings that some patients fear disclosing difficulties, it is possible that patients reported fewer difficulties than they would have done had the questionnaires been confidential to the researcher. However, as this study aimed to assess the feasibility of using the CORE-OM in secure hospitals the results could be seen to reflect patient responses in real clinical settings. While no significant association was shown between social desirability and scores on the CORE-OM it may be that the MCSDS did not detect defensive responding in this sample. Further limitations relate to the qualitative analysis. While a quality audit was conducted by a member of the research team, an independent quality assessment was not conducted. Thus, while the a priori themes were based on previous research it must be acknowledged that the sub-themes and levels are influenced by the perspective of the researcher.

Difficulties with the sample should be acknowledged. Although basic demographic data was obtained, no data was collected on educational history, marital status, or offending history preventing a full exploration of patient characteristics and any potential relationship with CORE-OM scores. Participants were excluded if they were not fluent in the English language and, as such, the majority (88%) of participants are ‘White British’ meaning the results cannot be generalised to those of other ethnicities. Due to difficulties with recruitment, and a potential difference in responses, women were excluded from this study. Thus, the findings are not generalisable to women in secure hospitals and further research is needed to determine the feasibility of using the CORE-OM with this client group.

Additional measures could have been used within this study. When investigating response styles malingering was assessed by determining whether patient responses on the CORE-OM were significantly higher than patient scores and it may have been preferable to have included a formal measure of malingering for statistical comparison. While the MIRECC GAF was used as a comparator to the CORE-OM the research may have benefitted from the use of other measures for anxiety, depression, trauma and wellbeing to assess criterion validity. However, it should be noted that the aim of this research was to assess feasibility of using the CORE-OM in secure settings rather than to determine validity and, as such, the decision was made to sacrifice the use of comparator measures in order to reduce the strain on the patient thereby increasing participation. Having completed the first stage further research should now focus on determining reliability and validity of the CORE-OM in forensic settings.

5.6. Further research

As this study aimed to be the first stage in accessing the feasibility of using the CORE-OM with forensic patients there was always the need to follow it up with further research investigating the reliability and validity using a larger sample size. Risk scores in particular were surprising low leading to the need to investigate this further, potentially through the use of formal risk assessments, by removing the time restriction on the CORE-OM or by asking staff to rate potential risk outside of the hospital setting. Here it may be useful to compare the staff scores on the CORE-OM with their ratings on the HONOS-secure to determine if the CORE-OM is able to identify relevant risk issues.

There is the need to conduct a feasibility study with a variety of other forensic patients including those with a diagnosis of personality disorder, those of different ethnicities, and women. The results suggest that nursing style may impact on their perception of patient difficulties and this needs further research. Further research could also focus on whether correlations improve if measures are dichotomous.

With patients reporting a fear of disclosing difficulties there is the need for further research, investigating differences in scores on a measure when completed by a key nurse and when completed anonymously. This area would also benefit from further qualitative exploration.

6. Conclusion

In summary, while the CORE-OM (PV) meets the criteria for use in secure hospitals, the low scores are incongruent with the *raison d'etre* for patients being in the secure settings and further research is needed to explore this phenomenon. Furthermore, associations between patient and nurse ratings on the CORE-OM are weak in all areas aside from functioning. Further research is needed to determine the validity of using the self-report CORE-OM in secure hospitals and this research will be carried forward in an ongoing programme of work.

References

Andrews, P. & Mayer, R.G. (2003). Marlowe-Crowne Social Desirability Scale and Short Form C: Forensic Norms. *Journal of Clinical Psychology, 59,* 483-492.

Atkisson, C., Cook, J., & Karno, M. (1992). Clinical services research. *Schizophrenia Bulletin, 18,* 561-626.

Baer, R.A., & Miller, J. (2002). Underreporting of psychopathology on the MMPI-2: a meta-analytic review. *Psychological Assessment, 14,* 16–26.

Bagby, R.M., Rogers, R., & Buis, T. (1994). Detecting malingered and defensive responding on the MMPI-2 in a forensic inpatient sample. *Journal of Personality Assessment, 62,* 191–203.

Barkham, M., Gilbert, N., Connell, J., Marshall. C., & Twigg, E. (2005). Suitability and utility of the CORE-OM and CORE-A for assessing severity of presenting problems in psychological therapy services based in primary and secondary care. *British Journal of Psychiatry, 186*, 239-246.

Barkham, M., Margison, F., Leach, C., Lucock, M., Mellor-Clark, J., Evans, C., et al. (2001). Service profiling and outcomes benchmarking using the CORE-OM: Towards practice-based evidence in the psychological therapies. *Journal of Consulting and Clinical Psychology, 69,* 184-196.

Barkham, M., Mullin, T., Leach, C., Stiles, W. B., & Lucock, M. (2007). Stability of the CORE-OM and the BDI-I prior to therapy: Evidence from routine practice *Psychology and Psychotherapy: Theory, Research and Practice, 80,* 269-278.

Beauford, J.E., McNiel, D.E,, & Binder, R.L. (1997). Utility of the initial therapeutic alliance in evaluating psychiatric patients' risk of violence. *American Journal of Psychiatry, 154,* 1272– 1276.

Beck, A.T., Ward, C.H., Mendelson, M., Mock, J., & Erbaugh, J. (1961). An inventory for measuring depression. *Archives of General Psychiatry, 4,* 561–571.

Bower, P., Gilbody, S., & Barkham, M. (2006). Making decisions about patient progress: the application of routine outcome measurement in stepped care psychological therapy services. *Primary Care Mental Health, 4,* 21-28.

Bowring-Lossock, E. (2006). The forensic mental health nurse: a literature review. *Journal of Psychiatric and Mental Health Nursing,* 13, 780-785.

Callaly, T. & Hallebone, E.L. (2001). Introducing the routine use of outcomes measurement to mental health services. *Australian Health Review, 24,* 43-50.

Cohen, A. & Eastman, N. (1997). Needs assessment for mentally disordered offenders and other requiring similar services: Theoretical issues and a methodological framework. *British Journal of Psychiatry*. *171*, 412-416.

Crabtree, B.F. & Miller, W.L. (1999). Using codes and code manuals: a template organizing style of interpretation. In Crabtree, B.F. and Miller, W.L. (Eds.), *Doing qualitative research.* 2nd Edition. Thousand Oaks, California: Sage Publications.

Cripps, J., Duffield, G., & James, D. (1995). Bridging the gap in secure provision: evaluation of a new local combined locked forensic/intensive care unit. *The Journal of Forensic Psychiatry, 6,* 77-91.

Cure, S., Chua, W.L., Duggan, L., & Adams, C. (2005). Randomised controlled trials relevant to aggressive and violent people, 1955–2000: a survey. *British Journal of Psychiatry, 186,* 185-189.

Dale, C., & Storey, L. (2004). High, medium and low security care: does the type of care make any difference to the role of the forensic mental health nurse. *Nursing Times Research, 9,* 168.

Department of Health. (2004). *Organising and delivering psychological therapies*. The Stationary Office: London.

Department of Health. (2008). *High quality care for all: NHS next stage review - Final report*. The Stationary Office: London.

Department of Health. (2010). *Equity and excellence: Liberating the NHS*. The Stationary Office: London.

Derogatis, L.R., Lipman, R.S., & Covi, L. (1973). SCL-90: an outpatient psychiatric rating scale – preliminary report. *Psychopharmacology Bulletin, 9,* 13-28.

Deshields, T.L., Tait, R.C., Gfeller, J.D., & Chibnall, J.T. (1995). Relationship between social desirability and self-report in chronic pain patients. *Clinical Journal of Pain*, 11, 189-193.

Eisen,S.V., Dickey, B., & Sederer, L.I. (2000). A self-report symptom and problem rating scale to increase patient involvement in treatment. *Psychiatric Services 51,* 349 -353.

Evans, C., Connell, J., Barkham, M., Margison, F., McGrath, G., Mellor-Clark, J., & Audin, K. (2002). Towards a standardised brief outcome measure: psychometric properties and utility. *British Journal of Psychiatry, 180,* 51-60.

Farrell, G.A. (1991). How accurately to nurses perceive patients needs? A comparison of general and psychiatric settings. *Journal of Advanced Nursing, 16,*1062-1070.

Faul, F. & Erdfleder, E. (1992). *GPOWER: A priori, post hoc and compromise power analyses for MS-DOS (computer program).* Bonn, FRG: Bonn University, Dept of Psychology

Fisher, R.A. (1915). Frequency distribution of the values of the correlation coefficient in samples of an indefinitely large population. *Biometrika, 10,* 507-521.

Fitzpatrick, E., 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,* 9-27.

Fonagy, P., Matthews, R., & Pilling, S. (2004). *The Mental Health Outcomes Measurement Initiative: Report from the Chair of the Outcomes Reference Group*. National Collaborating Centre for Mental Health: London.

Gierok, S.D., Dickson, A.L., & Cole, J.A. (2005). Performance of forensic and non-forensic adult psychiatric inpatients on the test of memory malingering. *Archives of Clinical Neuropsychology, 20, 755–760.*

Gildberg, F.A, Elverdam B. & Hounsguaard, L. (2010). Forensic psychiatric nursing: a literature review and thematic analysis of staff-patient interactions. *Journal of Psychiatric and Mental Health Nursing, 17,* 359-368.

Hawes, S.W. & Boccaccini, M.T. (2009). Detection of over-reporting of psychopathy on the personality inventory: a meta-analytic review. *Psychological Assessment, 21,* 112-124.

Heilbrun, K., Bennett, W. S., White, A. J., & Kelly, J. (1990). An MMPI-based empirical model of malingering and deception. *Behavioral Sciences & the Law, 8,* 45-53.

Heilbrun, K. (1992). The role of psychological testing in forensic assessment. *Law and Human Behaviour, 16,* 257-272.

Helmes, E. (2009). Conventional and actuarial methods to detect response distortion on the basic personality inventory. *Canadian Journal of Behavioural Science, 41,* 121-132.

Hopko, D. R , Averill, P. M. , Cowan, K., & Shah, N. (2002) Self-reported symptoms and treatment outcome among non-offending involuntary inpatients. *Journal of Forensic Psychiatry & Psychology, 13,* 88-106.

Horvath A.O, & Symonds B.D. (1991). Relation between working alliance and outcome in psychotherapy: a meta-analysis. *Journal of Counseling Psychology, 38,* 139–149.

Hunter, R., McClean,J., Peck, D., Pullen, I., Greenfield, A., Mcarthur, W., Quinn, C., Eaglesham, J., Hagen., & Norrie, J. (2004). The Scottish 700 outcomes study: A comparative evaluation of the Health of the Nation Outcome Scale (HoNOS), the Avon Mental Health Measure (AVON), and an Idiographic (OPUS) in adult mental. *Journal of Mental Health, 13,* 93-105.

Jacobs, R. (2007). *Investigating Patient Outcome Measures in Mental Health: Research Report for the OHE Commission on NHS Productivity*. Centre of Health Economics, York.

Kent, G. (2000). Understanding the experiences of people with disfigurements: An integration of four models of social and psychological functioning. *Psychology, Health & Medicine,* **5,** 117-129.

King, N. (1998). Template Analysis. In G. Symon & C. Cassel C., (Eds.), *Qualitative methods and analysis in organisational research: A practical guide.* London: Sage. Publications.

King, N. (2004). Using templates in the thematic analysis of text. In C. Cassell & G. Symon (Eds.), *Essential guide to qualitative methods in organizational research*. London: Sage Publications.

King, N., Thomas, K., Bell, D., & Bowes, N. (2003). *Evaluation of the Calderdale and Kirklees out of hours protocol for palliative care: Final report*. Primary Care Research Group, School of Human and Health Sciences, University of Huddersfield

Leach, C., Lucock, M., Barkham, M., Noble, R., Clarke, L., & Iveson, S. (2005). Assessing risk and emotional disturbance using the CORE-OM & HoNOS outcome measures. *Psychiatric Bulletin, 29,* 419-422.

Leach, C., Lucock, M., Barkham, M., Noble, R., & Iveson, S. (2006). Transforming between Beck Depression Inventory and CORE-OM scores in routine practice. British *Journal of Clinical Psychology, 45,* 153-166.

Lindemayer, J.P, Kay, S.R., & Plutchik, R (1992). Multivantaged assessment of depression in schizophrenia. *Psychiatry Research, 42,* 199-207.

Mann, S.J., & James, G.D. (1998). Defensiveness and essential hypertension. *Journal of Psychomatic Research, 45,* 139–148.

Martin, T., & Street, A.F (2003). Exploring evidence of the therapeutic relationship in forensic psychiatric nursing. *Journal of Psychiatric and Mental Health Nursing, 10,* 543-551.

McClintock, T. & Evans, C. (1995). Admissions to a close supervision unit: do patients with short admissions constitute the revolving door patients of minimum secure provision? *The Journal of Forensic Psychiatry, 6,* 63-75.

McCloskey, H. (2001). Evaluation of the CORE Outcome Measure in a therapeutic forensic setting. *The British Journal of Forensic Practice. 3,* 22-28.

McEvoy J.P, Apperson, I.J, Appelbaum, P.S. et al. (1989). Insight in schizophrenia: Its relationship to acute

Morlan, K.K., & Tan, S.Y (1998) Comparison of the Brief Rating Scale and the Symptom Inventory. *Journal of Clinical Psychology. 54*.7. 885-894.

Murray, K. (1996). The use of beds in the NHS medium secure units in England. *TheJournal of Forensic Psychiatry*. (3) 504-524.

Najim, H. & McCrone, P. (2005). The Camberwell Assessment of Need: comparison of assessments by staff and patients in an inner-city and semi-rural community area. *Psychiatric Bulletin, 29*, 13-17.

Niv, N., Cohen, A.N., Sullivan, G., & Young, A.S. (2007). The MIRECC Version of the Global Assessment of Functioning Scale: Reliability and validity. *Psychiatric Services, 58,* 529-536.

Nolan, P. (1989). Face value. *Nursing Times, 85,*62-65.

Office of Health Economics (2008). *Report of the Office of Health Economics on NHS performance and productivity*. Office of Health Economics, York.

Overall, J.E., & Gorham, D.R. (1962). The Brief Psychiatric Rating Scale. *Psychological Reports, 10,* 799-812.

Parker, J. & Gardner, G. (1992). The silence and the silencing of the nurse's voice: a reading of patient progress notes. *Australian Journal of Advanced Nursing, 9,* 3-9.

Preston, N.J & Harrison, T.J (2003). The Brief Symptom Inventory and the Positive and Negative Syndrome Scale: Discriminant validity between a self-reported and observational measure of pathology. *Comprehensive Psychiatry, 44*, 220-226.

Rask, M. & Lavander, S. (2000). Interventions in the nurse-patient relationship in forensic psychiatric nursing care: a Swedish study. *Journal of Psychiatric and Mental Health Nursing, 8,* 323-333.

Reynolds, W.M. (1982). Development of reliable and valid short forms of the Marlowe–Crowne Social Desirability Scale. *Journal of Clinical Psychology, 38,* 119–125.

Rogers, R. (1984). Towards an empirical model of malingering and deception. *Behavioural Sciences & the Law, 2,* 93-111.

Rogers, R. (Ed.) (1988). *Clinical assessment of malingering and deception*. New York: Guilford.

Sederer, L.I., Dickey, B., Eisen, S.V. (1997). Assessing outcomes in clinical practice. *Psychiatric Quarterly, 68,* 311–325.

Selten, J.P., Wiersma, D., van den Bosch, R.J. (2000). Clinical predictors of discrepancy between self-ratings and examiner ratings for negative symptoms. *Comprehensive Psychiatry, 41*, 191–196.

Shapiro, S.S., & Wilk, M.B. (1965). An analysis of variance test for normality (complete samples). *Biometrika, 52*, 591–611.

Slade, M. (1994). Needs assessment: Involvement of staff and users will help to meet needs. *British Journal of Psychiatry, 165*, 293-296.

Smith, J., Parker, J., & Donovan, M. (1991). Female admissions to a regional secure unit. *Journal of Forensic Psychiatrry, 21*, 95-102.

Swinton, M., Maden, A., & Gunn, J (1994). Psychiatric disorder in life-sentenced prisoners. *Criminal Behaviour and Mental Health, 4*, 10-20.

Taylor, P. (1998). Patients as intimate partners: resolving a policy crisis. In C. Kaye & A. Francey (Eds.), Managing high security psychiatric care. London: Jessica Kingley.

Waring, T., & Wainwright, D. (2008). Issues and challenges in the use of Template Analysis: Two comparative case studies from the field. *The Electronic Journal of Business Research* Methods, 6, 85-94.

Wing, J.K., Curtis, R.H. & Beevor, A.S. (1996). *HoNOS: Health of the Nation Outcome Scales: Report on Research and Development July 1993-December 1995*. London: Royal College of Psychiatrists.

Whewell, P. & Bonanno, D. (2000). The Care Programme Approach and risk assessment of borderline personality disorder: Clinical validation of the CORE risk subscale. *Psychiatric Bulletin, 24*, 381-384.

Section One: Overall Appendices

1.5. Ethical Approval

North Nottinghamshire Research Ethics Committee

Dear Ms Perry,

Study Title:	**Determining the Psychometric Properties and Utility of the CORE-OM in Forensic Hospitals.**
REC reference number:	**09/H0407/36**
Protocol number:	**2**

Thank you for your letter of 20 August 2009, responding to the Committee's request for further information on the above research and submitting revised documentation.The further information has been considered on behalf of the Committee by the Vice-Chair.

Confirmation of ethical opinion

On behalf of the Committee, I am pleased to confirm a favourable ethical opinion for the above research on the basis described in the application form, protocol and supporting documentation subject to the conditions specified below.

Ethical review of research sites

The favourable opinion applies to all NHS sites taking part in the study, subject to management permission being obtained from the NHS/HSC R&D office prior to the start of the study

Conditions of the favourable opinion

The favourable opinion is subject to the following conditions being met prior to the start of the study. Management permission or approval must be obtained from each host organisation prior to the start of the study at the site concerned.

For NHS research sites only, management permission for research ("R&D approval") should be obtained from the relevant care organisation(s) in accordance with NHS research governance arrangements. **It is the responsibility of the sponsor to ensure that all the conditions are complied with before the start of the study or its initiation at a particular site (as applicable).**

Statement of compliance

The Committee is constituted in accordance with the Governance Arrangements for Research Ethics Committees (July 2001) and complies fully with the Standard Operating Procedures for Research Ethics Committees in the UK.

Yours sincerely,

Mr Robert Johnson / Ms Trish Wheat

Vice-Chair / Committee Coordinator

1.6. Journal Approval

REMOVED

1.7. Journal Guidance: *Psychology and Psychotherapy*: *Theory, Research and Practice*

REMOVED

1.8. Journal Guidance: *The Journal of Mental Health*

REMOVED

Section Two: Literature Review Appendices

2.1.Table: Quality assessment for the reviewed papers

First Author & Year	**Craig 1976**	**Addington 1992**	**Blanchard 1992**	**Lindenmayer 1992**	**Addington 1993**	**Hamera 1996**
Quality Checklist Criteria						
Objective clearly described	1	1	1	1	1	1
Main outcomes described	1	1	1	1	1	1
Participant characteristics described	1	1	1	1	1	1
Distribution of confounders described	0	0	0	1	1	0
Main findings described	1	1	1	1	1	1
Estimates of random variability	1	1	1	0	1	1
Patients lost to follow up described	1	1	1	1	1	1
Actual probability values reported	0	0	0	0	0	0
Approached sample representative of whole population	1	1	1	0	1	1
Consenting sample representative of approached sample	0	0	0	1	0	0
Representative of real world services	1	1	1	1	1	1
Data-dredging described	1	1	1	1	1	1
Appropriate statistical tests	1	1	1	1	1	1
Valid and Reliable Measures	1	1	1	1	1	1
Same period of time for recruitment	1	1	1	1	1	1

Adjustment for cofounding variables	0	0	1	1	1	1
Sufficient sample size	0	1	1	1	1	1
TOTAL Quality Checklist Score (Total Possible Score = 17)	**12**	**13**	**14**	**14**	**15**	**14**

Author & Year	**Craig 1976**	**Addington 1992**	**Blanchard 1992**	**Lindenmayer 1992**	**Addington 1993**
COSMIN Criteria					
Content validity (construct)	1	1	1	1	1
Content validity (population)	1	1	1	1	1
Hypothesis Testing - Direction of predicted correlation described	1	0	0	1	1
Convergent Validity (description)	1	1	1	1	1
Convergent validity (measurement properties)	1	1	1	1	1
Criterion validity (gold standard/reasonable criterion)	1	1	1	1	1
Interpretability (distribution of scores described)	0	1	0	1	1
Generalisability (Missing responses)	0	0	1	0	1
Sample too small?	**Yes (0)**	No (1)	No (1)	No (1)	No (1)
COSMIN Rating	**Poor (6)**	**Fair (7)**	**Fair (7)**	**Good (8)**	**Excellent (9)**
Total Quality and COSMIN Rating (Total Possible = 26)	**18** **Excluded**	**20**	**21**	**23**	**24**

First Author & Year	**Morlan 1998**	**Huppert 2002**	**Preston 2003**	**Hunter 2004**	**Bell 2006**
Quality Checklist Criteria					
Objective clearly described	1	1	1	1	1
Main outcomes described	1	1	1	1	1
Participant characteristics described	1	1	1	1	1
Distribution of confounders described	0	1	0	1	1
Main findings described	1	1	1	1	1
Estimates of random variability	1	1	1	0	1
Patients lost to follow up described	1	1	1	1	1
Actual probability values reported	0	0	0	0	0
Approached sample representative of whole population	1	1	1	1	1
Consenting sample representative of approached sample	0	0	0	1	0
Representative of real world services	1	1	1	1	1

Data-dredging described	1	1	1	1	1
Appropriate statistical tests	1	1	1	1	1
Valid and Reliable Measures	1	1	1	1	1
Same period of time for recruitment	1	1	1	1	0
Adjustment for cofounding variables	1	0	1	0	1
Sufficient sample size	1	1	1	1	1
TOTAL Quality Checklist Score	**14**	**14**	**14**	**14**	**14**

Author & Year	**Morlan 1998**	**Huppert 2002**	**Preston 2003**	**Hunter 2004**	**Bell 2006**
COSMIN Criteria					
Content validity (construct)	1	1	1	1	1
Content validity (population)	1	1	1	1	1
Hypothesis Testing - Direction of predicted correlation described	0	0	0	1	0
Convergent Validity (description)	1	1	1	1	0
Convergent validity (measurement properties)	1	1	1	1	0
Criterion validity (gold standard/reasonable criterion)	1	1	1	1	1
Interpretability (distribution of scores described)	1	1	1	0	1
Generalisability (Missing responses)	1	0	0	1	0
Sample to small?	No (1)	No (1)	No (1)	No (1)	No (1)
COSMIN Rating	**Good (8)**	**Fair (7)**	**Fair (7)**	**Good (8)**	**Fair (5)**
Total Quality and COSMIN Rating (Total possible = 26)	**22**	**21**	**21**	**22**	**19**

2.2 Table: Design, statistical analysis and power of reviewed papers

First Author	Design	Statistical Analysis 2	Power 3	Inter-rater reliability (ICC)
Addington 1993	Between	Pearson's Product Moment Correlation	r = 0.69, Alpha = 0.05, Sample = 150, Power = 1.00 *Post hoc*	.89
Lindenmayer 1992	Between	Correlations (unspecified)	r = 0.58, alpha = 0.05, sample = 26, Power = 0.97 *Post hoc*	Not reported
Morlan 1998	Between	Pearson's Product Moment Correlation	***Reported*** r = 0.3, Alpha = 0.05, Sample = 27, Power = 0.83	Not reported
Hunter 2004	Between and Within	Kappa statistics	r = 0.4, Alpha = 0.05, Sample = 695, Power = 1.00 *Post hoc*	Not reported
Blanchard 1992	Between and Within	Pearson's Product Moment Correlation	r = 0.56, Alpha = 0.05, Sample = 30, Power = 0.97 *Post hoc*	.67 to .91
Hamera 1996	Between	Spearman's Rank Order Correlations	r = 0.66, Alpha = 0.05, Sample = 29, Power = 0.99 *Post hoc*	.78
Huppert 2002	Between	Correlations (unspecified)	r = 0.55, Alpha = 0.05, Sample = 33, Power = 0.98 *Post hoc*	Not conducted (although researcher trained in measures)
Preston 2003	Between and Within	Pearson's Product Moment Correlation	r = 0.46, Alpha = 0.05, Sample = 69, Power = 0.99 *Post hoc*	Not reported

Addington 1992	Between and Within	Pearson's Product Moment Correlation	r = 0.73, Alpha = 0.05, Sample = 150, Power = 1.00 *Post hoc*	CDS .89 HDRS .93
Bell 2006	Between and Within	Pearson's Product Moment Correlation	r = 0.31 Alpha = 0.05 Sample = 273 Power = 0.99	Not reported

2.3 Table: A Table of Acronyms for all measures discussed in the papers reviewed

Table: A Table of acronyms for all measures reported in the reviewed papers

Acronym	**Full Name of Measure**
ADIS- GAD	Anxiety Disorders Interview Schedule for DSM-IV schedule for Generalised Anxiety Disorder
ADIS – OCD	Anxiety Disorders Interview Schedule for DSM-IV schedule for Obsessive Compulsive Disorder
ADIS – PD	Anxiety Disorders Interview Schedule for DSM-IV schedule for Panic Disorder
ADIS – SP	Anxiety Disorders Interview Schedule for DSM-IV schedule for Social Phobia
ADIS-MD	Anxiety Disorders Interview Schedule for DSM-IV schedule for Major Depression
ASI	Anxiety Sensitivity Index
AVON	Avon Mental Health Measure
BDI	Beck Depression Inventory
BPRS	Brief Psychiatric Rating Scale
BSI	Brief Symptom Inventory
CDS	Calgary Depression Scale
DASS	Depression, Anxiety and Stress Scale
EPQ	Eyesenck Personality Questionnaire
HDS	Hamilton Depression Scale
HONOS	Health of the Nations Outcome Scale
MS	The Mood Scale
NEO	Neo Five Factor Inventory
NOSIE	Nurses Observation Scale for Inpatient Assessment
OCI	Obsessive Compulsive Inventory
PANSS	Positive and Negative Syndrome Scale
PPI	Personality Profile Index
PSI	Psychiatric Symptom Inventory
PSWQ	Penn State Worry Questionnaire
RDS	Raskin Depression Scale
SIAS	Social Interaction Anxiety Scale
SCL-90	Symptom Checklist-90
SDQ	Symptom Distress Questionnaire
SIQ	Symptom Intensity Questionnaire
SPS	Social Phobia Scale
SUMD	Scale to Assess Unawareness of Mental Disorder

Section Three: Research Report Appendices

3.6. Measures developed for the research

Blank for CORE-OM SV

The Brief CORE-OM Interview – Patient Version (Perry, Barkham and Evans., 2010)

Please use the blank copy of the CORE-OM if you need it.

1. When we filled out this questionnaire, what did you think about it?

2. Were there any questions which you did/did not like and why?

3. Would you prefer to complete the questionnaire on your own or with a key worker?

4. Do you think it would be useful for a member of your clinical team to use with you?

5. Do you think it would have been useful if a member of your clinical team had used this questionnaire with you at admission?

6. Did you feel you were able to be open in your responses?

7. Did you worry that answering this questionnaire would affect your treatment or make any difference to how long you stay in hospital?

8. Do you feel you have a good enough understanding of your feelings to answer the questions?

9. Do you have any other comments about the questionnaire?

The Brief CORE-OM Interview – Staff Version (Perry, Barkham & Evans., 2010)

Please use the blank copy of the CORE-OM if you need it.

1. What was your overall opinion of using this questionnaire?

2. Do you think your patient could complete the questionnaire alone or with a key worker?

3. Do you think this measure would have any use across the service as a standard?

4. Do you think it is relevant to your patient group?

5. Do you think it would give you any information you did not already know (and so make a difference to care planning)?

6. Did you feel you had the knowledge/understanding of your patient to answer the questions?

7. Were any questions difficult to answer?

8. Did you feel under pressure to give the 'right' answer?

9. Do you have any other comments?

3.7. Patient published measures

Blank space for CORE-OM Pdf

INSIGHT & TREATMENT ATTITUDES QUESTIONNAIRE (ITAQ)

REMOVED

Marlowe- Crowne Social Desirability Scale – Short Form C (Reynolds, et al.)

REMOVED

3.8. Staff published measures

3.9. Final template – nurse interviews

High Order Theme	**Narrow Themes**	**Quotes (Total N = 24)**
1. Acceptability	**1.1 Structure**	**1.1.1 Simplicity (n = 6)** • “it’s quite simple” • “has good scaling of mood and behaviour” • “it’s very easy to answer” • “it’s a handy tool for patients” “ • they’re often frustrated at the length of some questionnaires and this isn’t too long” • “it covers as much as possible for a short questionnaire” **1.1.2 Difficulties with Categorising (n=8)** • “perhaps the questions are too simplistic and don’t allow for shades of difference and all the ways in which one can vary” • “some of them you have to think about” • “the staff questionnaire would need a ‘don’t know’ option” • “it’s all subjective and the 0-4 are difficult to answer on the criteria given” • “some of the categories don’t quite fit with the questions” • “it can be hard to be so precise in the timings” • “some were slightly ambiguous” • “some questions are too broad”
	1.2. Content	**1.2.1. Variety (n = 5)** • “it seems to cover quite a lot” • “it looks at everything from behaviour to feelings to functioning” • “It looks at a variety of problems like isolation” • “it would be useful to get a clear picture of peoples problems” • “it’s good because its psychologically and practically focussed unlike a lot of questionnaires” **1.2.2. Providing Comparisons (n = 3)** • “if you use it regularly you could use to chat about differences in presentation and investigate why they have changed” • “you could use it regularly to detect

		changes or use it as a baseline” - “it might be useful as a guide or tool for measuring progress” **1.2.3. Lack of Positivity (n = 3)** - “it could cause difficulties and flag up problems when there isn’t one - “…sometimes it felt like it pathologies things we all feel at times, for example fleeting thoughts of low mood or self harm” - “ It’s too focussed on negative aspects - if it flags up problems in this system and people are deemed suicidal imagine the problems that will cause”
	1.3. Accessibility to all Professions (n = 3)	- “It would be nice for us to just be able to grab something like this … I always have to think ‘what’s available?’ - it’s good to have a day to day measure which isn’t a specialist psychology measure … - “The more things like this the better .. a lot of these things are done by specialist services who come onto the wards and rarely see the patient”
2. Feasibility	**2.1. Feasible(n = 4)**	- “he could do it”(2x) - “he’s capable, although I don’t know if he’d do it” - “it might be easier for him to tick a box than say he has a problem”
	2.2. Emotional Support (n = 8)	- “he’d get too upset to do it alone - “ he’d just rip it up with frustration” - “he wouldn’t be able to cope with it” - “if people are very distressed they would struggle - “ it’s about supporting the individuals” - “If they came up with alarming answers you’d be able to support them” - “You need to be able to discuss difficulties with them”. - “he do it better if he could talk to me about anything that’s upsetting”
	2.3. Support with Skills (n = 8)	- “ he has difficulties with reading” (x3) - “he’d need help, he has reading and writing problems” - “His eyesight isn’t very good” - “they may struggle to understand and need it clarifying”

		• "His literacy skills aren't great" • "I think he'd get confused by some of the questions"
	2.4. Building Relationships (n = 4)	• "I think it would be useful to go through it with them to develop the relationship" • "useful as a first line aid to starting the therapeutic relationships" • "it's better to do it together to get to know each other" • "I'd rather do it with him so I can get to know him better"
3. Relevance	**3.1. Useful**	**3.1.1. sharing emotions (n = 6)** • "It seem pertinent to the patients feelings" • it gets a feel for their current mood " • "it focuses on feelings and it may lead them to be more open" • "yes it would help a patient express their feelings" • "it would give patients a chance to tell you if they're feeling low" • "it might help them to talk more openly" **3.1.2. Helping Nurses to structure sessions (n = 7)** • "It would give the named nurse a deeper insight into their patients". • "it might focus the nurse on the patients feelings" • "it made me think about what I don't talk to him about" • "it would reassure you that you'd asked everything" • "it's good because I don't always ask these kind of questions in sessions" • "it would help to focus named nurse sessions on a day to day basis and bring that focus onto the way the patient is feeling" • "it made me think about the things I ask him about in sessions"
	3.2. Not Useful	**3.2.1. Preference for talking (n = 5)** • I know him enough through discussions" • "sometimes the constant monitoring of patients means you are overburdened

		with tick boxes” • “I’d rather just talk to him” • “we have enough of these kind of things already” • “it doesn’t offer anything new” **3.2.2. Offending (n = 2)** • it didn’t cover the things I expected it to for those who have a history of violence” • It doesn’t have enough focus on offending **3.2.3 Psychosis (n = 3)** • “it may not pick up on important psychotic symptoms” • “there are other more useful measures for people with severe mental health problems” • “it seems more useful for those in a general mental health hospital … it doesn’t focus on severe mental health problems or personality disorder”
4. Suitability	**4.1. Patient Responding style**	**4.1.1.Malingering (n = 2)** • “some with change their real responses to get what they need or want” • “some might do it just to get attention” **4.1.2.Defensiveness (n = 5)** • “it relies on good engagement with the patient – are you getting the real answers or what they want you to hear …especially for those early on in their admission” • “They’d probably lie or mask their answers” • “ a key nurse could do it with them to challenge their responses” • “it depends on how honest the patient is” • “sometimes you ask how a patient is and they say ‘fine’ without pushing it you just accept this” **4.1.3.Insight (n = 3)** • “I think my patient doesn’t see his problems, he might say everything is fine” • “some have little self awareness and

		may not be able to answer it" • "it depends on their level of insight"
	4.2. It is Suitable (n = 14)	**Yes 14 (**quotes coded into other themes)
5. Treatment	**5.1. Yes** (it would provide **new information** and aid care planning)	**5.1.1. Raising New Issues (n = 5)** • "if something came up with serious implications it could be addressed" • "it might help to pull out issues you weren't aware of" • "possibly I probably wouldn't ask the same questions in a 1-1 session • "if they spoke about something you weren't aware of " • "you hope your patient will talk to you about everything but it's not always the case in these services" **5.1.2.To provide Structure (n = 3)** • "you would hope the issues had been highlighted anyway but its might serve as an aid • "it may help to identify a problem earlier" • "it might help give a rationale for some things" **5.1.3.Getting to know someone new (n = 3)** • "it might be helpful if you didn't already know the patient" • "it's probably more useful when you don't know the patient" • "it might give new staff a quick idea of the patient" **5.1.4. Aiding MDT work (n = 4)** • "it would help those who visit the patient from other services have an idea of what the patient is experiencing" • "it might help generate discussion amongst the MDT for things you had known individually but not collectively" • "each staff member could have different views on the patient • "it gives other something to have a look at

	5.2 No New Information (n = 9)	- “not with someone I’ve known this long” - “No ... my patient is good at explaining their feelings” - “no, it’s stuff you would pick up anyway” - “I would hope not but there is the possibility … this environment doesn’t always lead to the best therapeutic relationships” - “probably not” - “no” - “no, it feels like we already have everything covered” - I know how patients are without the structure of a questionnaire” - “I wouldn’t tell me anything didn’t already know”
6. Understanding	**6.1. Recognising Emotion**	**6.1.1.Yes (n =15)** **6.2.1. No - Recognising Emotion (n = 9)** - “it’s hard to know what someone is feeling” - “sometimes you think you know their feelings but you can’t be sure” - “the ones about his moods were hard to answer” - “how do I know if he’s happy – I can only comment on the behaviour shown” - “the ones about his feelings were hard because he doesn’t tell you” - “difficult to answer questions about feelings as they can be so changeable for him” - “it’s hard to answer questionnaires on how patients feel as I can only comment on their behaviour” - “ I can only go from what I see in presentations” - “I’m not sure I’ve ever got a full understanding of how he feels, I’d like to see more of him”
	6.2. Pressure to Understand	**6.2.1. No (n = 18)** - No, it’s just my job” (3x) - “no, assessments are a key part of my role” - “no I do this kind of thing everyday” - “ No I just went with my gut instinct” - “No, if I wasn’t sure I said sometimes”

		• “ No, I try to be as objective as I can be” • “Not really, I tried to stay in the middle with my answers when I wasn’t sure” • “No I want to get it right”. • “no – even if I got it wrong I’d still feel ok giving my opinion,” • “ a different member of staff might give different ratings but I have to give my own judgement” • “no I just give my honest answers” • “not at all” • “No, it gives me a chance to think – am I doing as good a job as I think I’m doing”. • “no, there is no point in feeling pressure” • No (3x) **6.2.2. Yes (n = 6)** • “yes as it can affect the treatment the patient gets so I want to make sure I’m getting it right” • “a little, it’s important that you judge it right” • “With research you do think ‘she’ll think I don’t even know him!’ • Yes, because his and my opinions on his mental health are often so different” • “in tribunals it can be pressurising but you can only be honest” • “you feel like you have to answer, even if the question doesn’t seem right

3.10. Final template – patient interviews

High Order Theme	Narrow Themes	Quotes (Total N = 34)
1. Acceptability	**1.1 Structure**	**1.1.1 Simplicity (n =9)** • The questions were easy for me to understand" • "it was pretty straightforward" • "they were easy to answer" • "It's good that it's the same questions for everyone ... makes it easier" • "it was easy to do ... I'm used to filling these kind of things out" • "they asked the right questions for the right things" • "the questions were pretty self explanatory" • "it felt easy enough" • "it was quick" **1.1.2 Difficulties with Categorising (n = 4)** • "sometimes times there weren't enough choices" • " it could be more detailed" • "sometimes it was hard to decide between sometimes and often" • "they were ok but some I wasn't really sure of the meaning"
	1.2. Content	**1.2.1. Comfortable (n = 9)** • "it felt comfortable answering them" • "sometimes these questions can be upsetting but these were ok" • "the questions were helpful" • "it seemed alright to give my answers" • "I've filled a lot of these kind of questionnaires out before and this seemed less intrusive than previous questionnaires" • "It was ok" (x2) • " none (of the questions) were too personal" • "it was ok to be asked the questions" **1.2.2. Interesting or useful (n = 6)** • "it's interesting, it gives an insight into an individual" • " the questions were good questions" • "it's good because it looks at how your mood has been lately" • "it was interesting, it jogged my memory about the past and where I

		have been" - "it was alright– it looked at all different parts of mental health - "I liked that it asked about lots of different things" **1.2.3 Potential emotional Impact (n = 3)** - "it was ok although some of the harder topics brought out a bit of grief" - "it made me more aware that my progress has been slow" - "sometimes it's hard to answer personal questions about what you think about yourself" **1.2.4 Lack of Positivity (n = 3)** - "They were ok, I was a bit disappointed that they were asking about depression etc - " there was no opportunity to really focus on the positive" - "it seemed quite focussed on self harm and people who would hurt themselves"
2. Feasibility	**2.1. Feasible (n = 8)**	- "I could do it on my own" (x6) - "I'd prefer to do it on my own because it's quite private" - "it's easier to tick a box than talk to people"
	2.2. Emotional Support (n = 10)	- "it would be better to do with someone, get to spend some time with staff" - "it'd be better to do with a key worker to talk about things" - I'd prefer it with someone who knows me and we could have a chat about it" - "with my named nurse, I could sit on my own but it would take ages and I'd get worked up doing it" - "I'd be happier doing it with someone else" - "with someone I'd known for a while to support me" (x2) - " I'd want to do it with my nurse - it might feel a bit intrusive if someone you didn't know gave it you" - "sometimes it can upsetting so it's easier to do it with someone you know" - " My key worker knows me really well and I trust him"

	2.3. Support with Skills (n = 13)	- "you need to do it with someone else to draw the detail out" - "I can't concentrate very well, so it would be better with someone else" - "Maybe with someone else to help explain some of it" (x2) - "I'd rather do it with a member of staff, my reading and writing aren't very good" (x4) - "I'd want someone with me, I'm terrible at filling these things out myself" - "I'd probably forget to do it" - "I'd find it easier to have someone read it out loud" - "you'd need someone to read it out to you" - "I'd prefer a staff member to be with me, I can't read well without my glasses"
	2.4. Building Relationships (n = 3)	- "I could do it on my own but it might be a good way to get to know people" - "if you do it with a key worker it would help you build a rapport" - "it would help your nurse to get to know you"
3. Relevance	**3.1. Useful**	**3.1.1. Sharing views (n = 9)** - "yes, it would be helpful when you want to put across your insight" - "the questions help make things clearer in your mind - "so the staff could see my points of view" (x3) - Yes, to find out about me" - "yes, it clarifies and puts into perspective your ways of thinking" - " it asks questions I wouldn't think about ....to discuss it with my nurse" - "it would be useful to do it regularly to show I'm changing" **3.1.2. Sharing emotions (n = 8)** - "yeah, the questions allow me to talk about it when I'm upset" - "yes, I could talk about how I'm feeling so they'd find out more about me" - "yes, I don't talk about my feelings a lot" (x"2) - "Yes, it looks at how I am more than other questionnaires"

		• "it would help me talk to my nurse" (x3) **3.1.3** – General ("it would be good", "it's ok for them to do")
	3.2. Not Useful	**3.2.1. lack of need (n = 6)** • "no, I can cope without the help of hospital" • "I don't need them" (x3) • "no, it wouldn't make any difference" (x2) **3.2.2. Preference for talking (n = 6)** • "no, it would be better to get to know me" (x2) • "I'd rather someone talked to me than did paperwork" • "I don't like questionnaires" • "no, it's too similar to the other questionnaires we fill out" • "no, we already do too much paperwork" **3.2.3. Individual Difference (n = 3)** • "I'm not sure of its usefulness because everyone is so different … different personality, different medications" • "I don't know if it would be enough for everyone" • "people are too different"
4. Suitability	**4.1. Admission**	**4.1.1. Distress at admission (n = 24)** • "I couldn't have done it when I first came in, my head had gone" • "I don't think people would do it when they are first admitted, they're too anxious" • "people would turn it down, you'd need to wait at least 6 months" • "no, not when I just came in .. the paranoia was too bad" • "it's ok now as it focuses on depression and anxiety but early on my problems were too big – it wouldn't have been relevant" • "no, I was too distressed" (x2) • " No, my scores would have been much higher at admission" • "no, things were too rough at admission" • "no, I was too unwell" (x4)

		- “people wouldn’t be able to do it” - “no it would have been too early for me to cope” - “I wouldn’t have been as honest, I’d have been frightened” - When people first come in there are paranoid and suspicious but answering these things feels ok once you are settled”. - “no I was too upset” - “you need time to settle in first” - “no the voices would have tried to stop me answering” - “I wouldn’t have been able to cope with it” - It’s too soon (x3) **4.1.2. Yes (n = 10)** - Yes (x4) - “I did have something similar” - “I’ve done them before” - “maybe, to get to know people” - “it would be good to get to know your key nurse” - “it could be used to see how you change from admission” - “maybe you could try”
5. Treatment	**5.1. Honest**	**5.1.1 Yes (34x)** **5.1.2. Additional comments (n = 2)** - “I have been today but that doesn’t mean I would in a regular nursing session, I might be worried about what I disclosed” - “you get used to talking to strangers so its normal”
	5.3 Feared Consequences	**5.2.1. Yes (n = 11)** - “a little bit” (x2) - “I know from the past that if I reveal too much the system around me changes” - I’ve learnt that in this system if you reveal things like depression you are punished for it – my room was stripped, I was placed on observations … I’ve learnt never to reveal those things in a place like this, I hope I can in the community”. - “I did before I started but they were pretty harmless questions”

		• “at first but the questions were ok” • “yes, the voices were telling me not to say too much but I ignored them” • “A lot of people don’t say what going on for fear of going on observations but I just say it” • “Other patients tell me - ‘don’t tell them!’” so you can worry” • “a bit, things change sometime when you talking about feeling down” • “I’d want to tell the truth, but they’d just put me on levels (observations)” **5.1.2.No (n = 21)** • “No” (x11) • “no, it wouldn’t make any difference to my care” • “no I’m here because of my index offence so this wouldn’t matter” • “no, the doctors don’t base their decisions on these things” • “no, my nurse knows about me anyway” • “my team know I’m depressed so it wouldn’t change anything” • “no, I haven’t got any difficulties so it won’t change anything” • “no -I’m ok” (x3) • “no, I’m ready to move on so I don’t want to keep anything private, I want to open”
6. Understanding	**6.1. Emotions**	**Yes (34x)** **Additional Comments (n = 2)** • “some of them but some not” • “yes, just about”