Examining forensic service effectiveness within an NHS Trust: Outcomes and considerations for practice

Nishant Krishnan
Swinburne University of Technology, Australia

Ivan Sebalo
University of New York in Prague, Prague

Wendy Copeland-Blair
NHS, England, UK

Jane L. Ireland[1]
School of Psychology and Humanities, University of Central Lancashire, UK

Disclosure Statement

The authors report there are no competing interests to declare.

[1] Corresponding author. Email: JLIreland1@uclan.ac.uk

Abstract

The current study examines whole service effectiveness using a secure version of the Health of the Nation Outcome Scales (HoNOS-Secure), further supplemented in some services by the Clinical Outcomes in Routine Evaluation – Outcome Measure (CORE-OM) and the Patient Reported Experience and Outcome Measure (PREOM). The utility of these measures were considered across the full remit of forensic services within a single NHS Trust. A total of 1,038 service users were included (male, n = 876), with the majority of these presenting with HoNOS-Secure ratings. It was predicted that there would be differences pre and post therapy indicated using these measures, that there would be further differences in relation to period of contact with services, and an association also noted in relation to aggression within services. Results indicated that HoNOS-Secure scores decreased following treatment and as a function of time spent in secure care, however no statistically significant improvement or deterioration were observed on patient-reported outcome measures (CORE-OM and PREOM). The paper concludes with a comparative examination evaluating potential arguments regarding why low levels of distress are observed within forensic populations, and why discrepancies exist between clinician-rated and patient-reported routine outcome measures.

Keywords: Service effectiveness; Forensic services; HoNOS; CORE-OM; PREOM

Introduction

Clinical outcomes and their management have, arguably, become essential as key performance indicators for services (Dickens et al., 2010), particularly since the emergence of increased public scrutiny and spending of public funds. Whereas previously, forensic services have almost exclusively relied upon measures of recidivism, readmission, or mortality (Coid et al., 2007) as indicators of performance, inpatient mental health settings have paradigmatically shifted away from independent, post-hoc indicators, towards the use of routinely administered outcome measures that capture 'whole service' effect. These measures, aside from establishing treatment effectiveness in real-life forensic settings, have been theorised to facilitate the multi-faceted evaluations of pharmacological interventions, psychological therapies and complex social interventions, which are all characteristics features of current inpatient mental health services (Ellwood, 1988). Given the significance placed on routinely administered outcome scales, evaluating the utility of such measures remains a continuous process of critical importance.

One of the most well-known and empirically supported routine outcome measure, the Health of the Nation Outcome Scales (i.e., HoNOS), is a clinician administered tool designed to monitor patient outcomes within mental health services (Wing et al., 1998). Since its inception in 1998, the HoNOS has come to be mandated in several international jurisdictions, including Australia, New Zealand (Shinkfield & Ogloff, 2016), and the United Kingdom (Dickens et al., 2010). Originally designed for working age adults (i.e., 18 to 65), the HoNOS comprises clinician-rated items relating to behaviour, impairment, symptoms, and social functioning outcomes (Dickens & O'Shea, 2017). However, despite the utility of the original HoNOS when used with civil populations (see Pirkis et al., 2005), numerous difficulties have been reported when applying the tool in specialist mental health settings, particularly forensic secure services (Shinkfield & Ogloff, 2016). The broad and disparate needs of forensic

patients as compared to civil populations (i.e., level of security, risk and risk management; Ogloff et al., 2015), together with the omission of a risk scale have raised concerns regarding the utility of the original HoNOS within secure services. Specifically developed to address these utility concerns, HoNOS-Secure, part of the broader HoNOS 'family of tools', is an adapted routine outcome measurement tool for use with forensic populations.

Expanded and refined over two iterations, HoNOS-Secure (previously referred to as HoNOS-MDO - mentally disordered offenders) provides a means of tracking the clinical, social and security needs of users of secure psychiatric services, prisons, and forensic community services (Sugarman et al., 2009). HoNOS-Secure contains items capturing clinical and social functioning and security. The latter monitors changes in a client's need for risk and security management procedures (Dickens et al., 2007). As the only truly mandated outcome measure for adults in the United Kingdom, the HoNOS 'family of tools' have been considered a complete, robust and rigorous measure of whole service effectiveness (Macdonald & Furgard, 2015).

Despite such assertions, in a forensic or secure services context, Ellwood's (1988) vision has had limited impact (Dickens & O'Shea, 2017, p. 162). Aside from criticisms being raised regarding the quality of data input – information has been found to be incomplete or missing across numerous cases and services (Delafon et al., 2012) – the utility of HoNOS-Secure as a research outcome measure has provided mixed results (Dickens & O'Shea, 2017). While some researchers have asserted that HoNOS-Secure is a reliable tool that can effectively track the needs of forensic mental health clients over extended periods of time (Burgess et al., 2006), others have argued that the scales are insufficiently sensitive to detect meaningful change in users of both inpatient and outpatient forensic services (Audin et al., 2001). The accurate classification of users on outcomes relevant to mental health care and containment of risk is an important function of any measurement tool (Pirkis et al., 2005),

particularly in the context of forensic or secure services, where decisions regarding leave, transfer, or release have serious and potentially significant consequences. Notwithstanding the significant efforts of researchers, questions regarding HoNOS-Secure's clinical utility continue to exist (Stein, 1999).

Routinely administered as a companion to the HoNOS-Secure, the Clinical Outcomes in Routine Evaluation – Outcome Measure (CORE-OM; Barkham et al., 2001) is a pan-theoretical patient reported measure of psychological distress (Barkham et al., 2005). The original CORE-OM consists of items that captures domains of subjective wellbeing, problems/symptoms, functioning, and risk to self or others. Noted both for its rigorous and iterative developmental process, the CORE-OM has been validated and deemed fit for use with community populations (Connell et al., 2007), older adults (Barkham et al., 2005), and primary and secondary care settings (Barkham et al., 2001). It has also been applied, although to a more limited degree, to forensic populations (e.g., Perry et al, 2013; McGrath et. al, (2020). More particularly, research has showed the tool is sensitive to both low and high intensity ranges of distress (Barkham et al., 2006), and therefore theoretically capable of reflecting the diversity in presentation of forensic service users. Despite these strengths, concerns have been raised regarding the degree to which patient reported measures, such as the CORE-OM, truly address the concerns of these individuals, particularly in domains relevant to 'recovery' as opposed to those considered of value to clinicians, managers, service providers, governments, and other key stakeholders (Macdonald & Furgard, 2015, p. 313).

In addition, neglecting the views of service users may provide a biased representation of whole service effect. Given this potential oversight, NHS England specialist commissioners introduced a requirement for low and medium secure services to utilise a Patient Reported Experience and Outcome Measure (PREOM) as a method of capturing issues of fundamental importance to service users and identifying potential service

improvement opportunities. However, to date, no research has sought to understand the extent to which the data captured through PREOM is of value in understanding progress, change and service effect. More broadly, there continues to be a paucity of research examining the clinical utility of HoNOS-Secure, CORE-OM, PREOM in forensic settings, particularly in relation to the CORE-OM and PREOM, with research also generally failing to capture whole service effectiveness (i.e., beyond one single site).

The current study aimed to provide some preliminary evidence of the utility, or otherwise, of the HoNOS-Secure, CORE-OM and PREOM in measuring patient clinical change, and whole service effectiveness within one Trust's forensic services. More particularly, the investigation aimed to explore any association with variables of service interest (i.e., therapy and aggression), and evaluate the tools value as an indicator of service effectiveness. It was hypothesised that: (1) HONOS-Secure ratings (i.e., clinician-rated outcomes relating to behaviour, impairment, symptoms, and social functioning) and CORE-OM ratings (psychological distress) would be higher pre-therapy compared to post-therapy, while PREOM ratings (patient's perceptions of their general wellbeing, care, safety and future) would be lower pre-therapy compared to post-therapy; (2) HoNOS-Secure and CORE-OM ratings would decrease over the service users' period of admission or contact with the relevant service; (3) Higher HoNOS-Secure ratings would be predictive of increased aggressive incidents relative to lower scores on the stated measures.

Method

Setting

The relevant Trust provides specialist secure mental health and learning disability care across eleven sites in England. Eligible patients were adults who, as a minimum, had two HoNOS-Secure, CORE-OM or PREOM assessments completed within the previous eight years (i.e., August 2014 to August 2022).

Patient demographics

There was a total of 1,038 patients who completed either a HoNOS-Secure, CORE-OM, or PREOM assessment between August 2014 and 2022. The majority of patients were male (n = 876), with a mean age of 36.29 years old (SD = 11.3) at first assessment. Twenty-one participants were missing accurate age-related data at time of first assessment. Most participants were of White background (83.14%), 5.88% from Black background, with 4.34% Mixed, 4.14% Asian, and 0.96% from Other backgrounds. Those failing to report ethnicity reached 1.16%. In terms of service classification and based on the information available, most were contained within High secure services (33.33%), followed by 22.64% from medium secure, and 13.29% from low secure units or various non-specific forensic services. A small fraction of participants were attached to forensic outreach services (0.67%), with the remainder having no specific or unclear recorded service level provision but were, neverthless, within forensic services.

Design

A retrospective study design was employed. As part of routine clinical practice, HoNOS-Secure, CORE-OM and PREOM assessments were completed by clinical teams. In August 2022, HoNOS-Secure, CORE-OM, PREOM assessments and demographic and clinical information was extracted from internal clinical patient systems as part of an approved service evaluation. This was a result of the requested information being routinely collected, provided in an anonymised format and used to evaluate an existing service. Consequently, it did not require approval from an NHS Research Ethics Committee. Rather, service evaluation approval was obtained from the Trust's Research and Development Department in December 2021 (SE2022-01).

Measures

Patient's age at assessment, gender, ethnicity, admission and discharge dates, and service type were extracted from patient records. In addition, information on incidents were obtained from records and the following data was extracted:

HoNOS-Secure. The HoNOS-Secure Version 2b comprises (a) modified versions of the original twelve HoNOS items, (1-12) that measure behaviour, impairment, symptoms, and social problems (Wing et al., 1998), and (b) seven security items (A-G) that track current clinical risk management needs (i.e., risk of harm to self or others, vulnerability, environment, staffing). Each item is rated on a five-point scale (0-4) with each point representing an anchor with an accompanying narrative description (Dickens & O'Shea, 2017). Evidence has shown that both HoNOS-Secure scales (i.e., 1-12, A-G) have (a) acceptable inter-rater reliability in routine clinical practice (Dickens et al., 2007), and (b) can detect statistically significant change over time when used to calculate a performance metric in patient cohorts (Dickens & O'Shea, 2017).

CORE-OM. The CORE-OM (Barkham et al., 2001) is a 34-item self-report measure developed to examine psychological distress. The measure comprises four domains, with each domain comprising specific clusters: subjective wellbeing, problems/symptoms, functioning, and risk to self or others. Each item is scored on a five-point scale from 0 to 4 (i.e., 'most or all of the time', 'often', 'sometimes', 'only occasionally', 'not at all'). Evidence has demonstrated good internal consistency (Cronbach's alpha=0.94) and test-retest reliability (Spearman's p=0.90; Evans et al., 2002).

PREOM. PREOM is a 10-item self-report measure designed to examine patient's perceptions of their general wellbeing, care, safety and future. The tool is intended to inform the care review process by aiding the formulation of an action plan and improving the overall care experience. Each item is scored on a 5-point scale, with higher scores indicating improved functioning and adjustment.

Data analysis

Data was analysed by fitting the information to the maximum likelihood (REML) linear mixed models with random slopes and intercepts. This method was used as it accounts for the variation within as well as between participant scores. Fixed effects represented the average difference between scores obtained at different time points, with earlier scores considered the intercept. Random effects reported in the tables show residual or unexplained variance. To ensure that mixed models were considered most appropriate, first, each model was fitted with a simple linear regression, followed by a random intercept, and then a random slope model. At each step, a Chi square test was used to determine whether a model with more elements provided a significantly better explanation of data variance. Random slope and random intercept models were chosen given data was obtained from the same participants, at different times. Consequently, each test operated with observations (i.e., data entries) as sample units rather than participants. For example, participants may have two HoNOS scores: one taken prior to the commencement of therapy, and one immediately post-intervention. Since there were sizeable variations in the number of assessments, each timepoint was used as a separate data entry.

Results

HoNOS-Secure, CORE-OM and PREOM all had good internal consistency (see Table 1). Given that not all participants were assessed with each of the scales, the values for first and last date of assessments are based on different samples, as outlined in each corresponding analysis.

<Insert Table 1 here>

Pre and post therapy: HONOS-Secure, CORE-OM and PREOM ratings

HoNOS-Secure. A total of 865 patients attended therapies *and* had corresponding HONOS-Secure values. The model for HoNOS-Secure security subscale with random slope and random

intercept was a better fit for the data than either the linear, Chi (1) = 506284.42, p = 0, or random intercept only models, Chi (2) = 42185.66, p = 0. Likewise, the model for HoNOS-Secure health and social subscale with random slope and random intercept was a better fit for the data than either the linear or random intercept only model, Chi (1) = 443110.79, p = 0 and Chi (2) = 50713.47, p = 0, respectively. On average, following the start of the therapy, scores on both the HONOS-Secure security subscale and health and social domains decreased (see Table 2). Thus, there was a reduction of security risks, and improvements in mental health domains following therapy.

CORE-OM. A total of 50 patients engaged in therapy *and* had corresponding values on CORE-OM. The model for total score on CORE-OM Wellbeing subscale with random slope and random intercept was a better fit for the data than either the linear or random intercept only model, Chi (1) = 10607.44, p = 0 and Chi (2) = 1521.53, p = 0, respectively. Similarly, the model for total score on CORE-OM Problems/Symptoms subscale with random slope and random intercept was a better fit for the data than either the linear and random intercept only model, Chi (1) = 8903.74, p = 0 and Chi (2) = 2305.08, p = 0, respectively. Likewise, the model for total score on CORE-OM Functioning subscale with random slope and random intercept was a better fit for the data than the linear and random intercept only model, Chi (1) = 9668.85, p = 0 and Chi (2) = 1141.6, p = 0, respectively. Further, the model for total score on CORE-OM Risk subscale with random slope and random intercept was a better fit for the data than either the linear and random intercept only model, Chi (1) = 12003.98, p = 0 and Chi (2) = 2798.91, p = 0, respectively. Finally, the models for total score on all CORE-OM subscales with random slope and random intercept was a better fit for the data than either the linear, Chi (1) = 9888.42, p = 0, or random intercept only model, Chi (2) = 2091.28, p = 0.

Across participants there were no differences on any CORE-OM total or subscale scores following engagement in therapy (see Table 2). In other words, from a patient's perspective, there was no significant improvement or deterioration on domains of wellbeing, problems, functioning or risk following treatment.

PREOM. A total of 109 patients attended therapy *and* had corresponding values on PREOM. The model for total score on PREOM with random slope and random intercept was a better fit for the data than either the linear or random intercept only model, Chi (1) = 17146.85, p = 0 and Chi (2) = 712.79, p = 0, respectively. However, there were no differences in PREOM total scores pre and post therapy, indicating that patients' subjective perception of well-being did not change post-treatment (see Table 2).

<Insert Table 2 here>

Service admission or contact time: HONOS-Secure and PREOM ratings

There were insufficient data points for CORE-OM (n = 19) for those assessed twice on this measure. Consequently, focus was on HoNOS-Secure and PREOM.

HoNOS-Secure. All models assessing change throughout hospital stay did not contain enough observations to use either random intercept or random slope models. Consequently, only models with a random intercept were tested. A total of 857 patients were in contact with services and had at least two assessments. The model for HoNOS-Secure with random intercept was a better fit for the data than the linear model, Chi (1) = 431.13, p = 0. Likewise, the model for HoNOS-Secure health and social domains subscale with random intercept was a better fit for the data than the linear model, Chi (1) = 185.24, p = 0. Across participants, scores on both secure and health and social subscales were lower at the last time of assessment relative to the first assessment timepoint (see Table 3). This indicates that patients risk as well as other needs improved as a function of time spent in the secure care.

PREOM. 109 patients were in contact with services and had corresponding values on PREOM. The model for total score on PREOM with random slope and random intercept was a better fit for the data than the linear model, Chi (1) = 23.14, p = 0. However, there were no differences in PREOM ratings from first to last date of assessment, indicating that patient's subjective perceptions of care neither improved nor deteriorated (see Table 3).

<Insert Table 3 here>

Aggression: HONOS-Secure

One hundred and six patients instigated an incident or were victimised and had corresponding values on HoNOS-Secure. The model for predicting the number of perpetrated incidents based on HONOS-Secure scores with random intercept was a better fit for the data than the linear model, Chi (1) = 3183.79, p = 0. Similarly, the model for predicting number of victimisation experiences by HONOS-Secure scores with random intercept was a better fit for the data than the linear model, Chi (1) = 3733.09, p = 0. Nevertheless, neither HoNOS-Secure subscales were significantly associated with the number of incidents perpetrated by patients or victimisation experiences. Thus, neither risk nor mental health needs were associated with acts of aggression or victimisation events. There was insufficient data to consider this in relation to the CORE-OM and PREOM.

Discussion

This study represents the first known effort at the empirical examination of whole service effectiveness across all Secure and Specialist Learning Disability Division sites within a single NHS Trust. As anticipated, findings confirmed stable improvement vis-à-vis the health, social and risk related needs of patients across diverse care pathways, providing preliminary evidence for the indication of Trust effectiveness in managing the needs of forensic service users. Despite these initial positive indications, a more complex picture emerged when scrutinising (a) the low levels of distress reported by the population, (b) the

clinical utility of clinician-rated compared to patient-reported routine outcome measures, and (c) value of HoNOS-Secure as a predictor of aggression and victimisation events in forensic populations.

Measuring service effectiveness and functional therapeutic change

Consistent with the prediction that HONOS-Secure ratings would be higher pre-therapy compared to post-therapy, scores on both HoNOS-Secure subscales decreased following the delivery of psychosocial treatment. Patients showed an improvement in functioning and a decrease in security related needs, indicating more broadly a reduced need for care and risk management intervention post-treatment. While these findings are promising and provide partial support for the continued use of Trust treatment programmes, evidential weight should be assigned cautiously. First, the prediction was only partially supported in that neither the PREOM or CORE-OM demonstrated change in the expected direction accounting for therapy, suggesting that the more global measure of impact (HONOS-Secure) was identifying change whereas patients did not. However, second, and perhaps more pertinently, the investigation did not specifically examine whether a certain therapy (e.g., CBT for psychosis) or class of treatments (e.g., violence reduction therapies) were effective. Rather, the investigation was at a collective level. It was impractical to examine effectiveness at an individual therapy level owing to the range of therapies provided (<50) and the data available for each therapy This precluded the drawing of any direct comparisons but highlights the need for additional research in this area.

Aside from functional change following treatment completion, consistent with the prediction was the finding that HoNOS-Secure ratings would decrease over the service users' period of admission or contact with the relevant service. More simply, patients showed an improvement in health, social and security related needs as a function of time spent in care of the Trust. This, again, was not replicated with the PREOM. However, in relation to the

HoNOS-Secure, it would appear the scores serve as a proxy measure of whole service effectiveness, providing collective support for the efficacy of psychosocial treatment programmes and management practices employed at Trust sites. While this is the first investigation to examine effectiveness across an entire NHS Trust, the findings are not unexpected given that individual investigations (at therapy level) have found treatment programmes and practices in place at the Trust are effective in reducing risk and security-related concerns (Daffern et al., 2018).

Clinician-rated versus patient-reported discrepancies on routine outcome measures

While pre- to post therapeutic and whole service improvements were expected on patient-reported outcome measures of CORE-OM and PREOM, statistically significant differences were not observed, as noted, when patients were asked to report their own perceptions of wellbeing following either treatment or generalised care. Though unanticipated, there are four potential explanations for these findings, which can be offered: (a) clinical justification, (b) sensitivity interpretation, (c) construct discrepancy, (d) patient and clinician biases. In addition, a further explanation is that these findings could represent an artefact of the available data; the PREOM and CORE-OM data was limited in relation to sample size (whereas the HONOS data was not) and there was no means by which we could ascertain if this data was truly representative of the patient experience.

The most clinically intuitive explanation, however, asserts that as a lack of insight regarding general wellbeing is a defining characteristic for those with a serious mental illness, learning disability and/or contained within secure services (Amador et al., 1993), patient-reported outcome measures may fail to accurately capture levels of distress and therefore, treatment improvement or deterioration. Evidence has emerged which indicates that forensic patients who experience severe and enduring mental health difficulties record mean total distress scores below clinical normative cut-offs (Gilling McIntosh, 2020). Low levels of

reported distress are at odds with the general presentation of the forensic population, with reasons for such underreporting as of yet unclear but a useful avenue for future research to pursue. What is accepted, however, is that self-report is determined by an individual's willingness and ability to accurately report their experiences and therefore reliant upon clinical insight (Bell et al., 2007). Impaired insight has been shown to influence several aspects of clinical outcomes, including treatment adherence and more pertinently, symptom severity (Bell et al., 2007). Thus, low levels of self-reported distress detected within the present study may indicate more broadly lack of patient clinical insight and could therefore explain why treatment or general improvements were not observed on patient-reported measures.

Notwithstanding the clinical justification, a more likely explanation could be the nature of the measures themselves. HoNOS-Secure relative to CORE-OM or PREOM was specially developed for secure populations and thus we would expect it to be more sensitive to and validated to identify change. A review of the evolution of the measure supports this assertion; HoNOS-Secure version 2 was advanced to compensate for the limitations of the original tool (Dickens et al., 2010), iteratively developed, and has attracted significant evidential backing from independent studies (Delaffon et al., 2012). In contrast to this, few studies have examined the utility of the CORE-OM outside of primary care settings and with forensic populations. Accordingly, any suggestion that the CORE-OM is of comparable sensitivity to HoNOS-Secure is likely overstated and may offer one explanation for the absence of prediction support.

Perhaps the most parsimonious explanation, the construct discrepancy interpretation claims that as HoNOS-Secure, CORE-OM and PREOM are overlapping rather than identical constructs. The measures capture different aspects of patient functioning and wellbeing. As identified by Delaffon et al. (2012), the correlation between HoNOS-Secure and CORE-OM

is “frequently low”, intimating at a dimensional discord between HoNOS-Secure and other patient-reported outcome measures. Delaffon et al’s. (2012) assertion together with the findings from the present study adds further weight to the supposition that forensic service users constitute a broad and disparate category of mental health patient (Shinkfield & Ogloff, 2015), whose experiences of secure services are not wholly or accurately captured by current patient-reported measures. In essence, we argue here that the measures may lack the required validity and/or specificity for this diverse population. Beyond these explanations, it is accepted that biased or inaccurate responding may account for null findings. Forensic patients are involuntarily detained, frequently presenting with severe and enduring mental illnesses, alongside chronic comorbid psychiatric disorders, neuropsychological impairments, and histories of trauma (Ogloff et al., 2015). Such adverse experiences encourage threat-based mentalities, often leading to a mistrust of authority (Perry, 2010). This distrust, together with a fear as to the consequences of honest disclosure, may lead to a minimising response style, skewing observations on measures such as the CORE-OM and PREOM (Perry, 2010).

Predicting aggression and victimisation

Finally, the intention of the authors was to investigate the predictive utility of all three routine outcome measures. However, due to insufficient instances of aggression and/or victimisation, such analyses was not possible and restricted only to HoNOS-Secure. As noted, findings in relation to aggressive incidents demonstrated no association between HoNOS-Secure and aggression/victimisation, with the prediction not therefore supported. The HoNOS-Secure was developed for the purposes of risk management rather than risk assessment to aid clinical decision-making rather than predicting violence, but predictive ability is comparatively limited (Finch et al., 2017, p. 412), with the current study raising further concerns in relation to the potential utility. Nevertheless, while HoNOS-Secure is not recommended for use *in place* of established risk assessment measures, the tool may hold some value in “informing

and directing risk assessment processes" (Finch et al., 2017, p. 424) owing to the noted evidence of the ability to identify change in relation to therapy and service admission time. Put simply, it may be premature to exclude the value of this measure without further research that perhaps aligns it more to a wider risk assessment battery, to determine if it has any additive value.

Limitations and future directions

As previously noted, HoNOS-Secure ratings were completed by clinicians during the course of routine clinical work. Although policy directives exist, ratings were not recorded at standard intervals and teams were not blinded to outcomes. In addition, there were variable sample sizes, which limited hypotheses testing, with no additional clinical measures routinely collected. This prevented comparisons between the examined measures and other indicators of risk or functioning. However, and as discussed by Dickens et al. (2010), these limitations "do reflect real-life practice and may be taken as an indication of the difficulties faced by clinical teams in detecting change" (p. 44). Future research could therefore explore how to acquire more useable data that accounts for service demands, leading to a more pragmatic solution to data collection. Such research could also explore the low levels of distress reported in forensic populations, specifically test the validity of the posited explanations (i.e., clinical justification, sensitivity interpretation, construct discrepancy and biases), as well as examining the utility of HoNOS-Secure, CORE-OM and PREOM using other indicators of risk and functioning.

Implications

There are several implications that could be drawn from the present findings, as follows;

- Incomplete datasets limited the types of analyses that can be conducted. Practitioners should endeavour to record comprehensive and accurate ratings at regular intervals, so as to allow for the examination of clinically meaningful change. A means of achieving

this could be via regular feedback at a service level of the outcome data so there is some clear gain and evidenced use of the data, making it more valuable for practitioners as a means of assessing the value of their interventions and contact with patients.

- Any estimation of whole service effectiveness is strengthened through an examination using complementary informational inputs. Therefore, services could support practitioners to collect a wider variety of routine outcome measures to accurately demonstrate efficacy to a greater degree.
- Responding biases may have skewed patient reports of functioning and wellbeing. The use of validity measures alongside the CORE-OM and PREOM is consequently recommended.
- Dual role biases may have unavoidably impacted on scores on clinician-rated measures. Services with sufficient capacity should ensure that clinician-rated measures are completed by practitioners outside of the direct treatment team, to afford a greater level of independence, where possible. This could, for example, include other members of the patient care team who know the client well.

References

Amador, X. F., Strauss, D. H., Yale, S. A., Flaum, M. M., Endicott, J., & Gorman, J. M. (1993). Assessment of insight in psychosis. The American Journal of Psychiatry, 150(6), 873-879. https://doi:10.1176/ajp.150.6.873

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

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 settings. *The British Journal of Psychiatry, 186*, 239-246. https://doi:10.1192/bjp.186.3.239

Barkham, M., Margison, F., Leach, C., Lucock, M., Mellor-Clark, J., Evans, C., . . . McGrath, G. (2001). Service profiling and outcomes benchmarking using the CORE-OM: toward practice-based evidence in the psychological therapies. Clinical Outcomes in Routine Evaluation-Outcome Measures. *Journal of Consulting and Clinical Psychology, 69*(2), 184-196. https://doi:10.1037//0022-006x.69.2.184

Bell, M., Fiszdon, J., Richardson, R., Lysaker, P., & Bryson, G. (2007). Are self-reports valid for schizophrenia patients with poor insight? Relationship of unawareness of illness to psychological self-report instruments. *Psychiatry Research, 151*(1-2), 37-46. https://doi:10.1016/j.psychres.2006.04.012

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

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

Connell, J., Barkham, M., Stiles, W. B., Twigg, E., Singleton, N., Evans, O., & Miles, J. N. (2007). Distribution of CORE-OM scores in a general population, clinical cut-off points and comparison with the CIS-R. *The British Journal of Psychiatry, 190*, 69-74. https://doi:10.1192/bjp.bp.105.017657

Daffern, M., Simpson, K., Ainslie, H., & Chu, S. (2018). The impact of an intensive inpatient violent offender treatment programme on intermediary treatment targets, violence risk and aggressive behaviour in a sample of mentally disordered offenders. *Journal of*

Forensic Psychiatry & Psychology, 29(2), 163–188. https://doi.org/10.1080/14789949.2017.1352014

Delaffon, V., Anwar, Z., Noushad, F., Ahmed, A. S., & Brugha, T. S. (2012). Use of Health of the Nation Outcome Scales in psychiatry. *Advances in Psychiatric Treatment, 18*(3), 173-179. https://doi:10.1192/apt.bp.110.008029

Dickens, G. L., & O'Shea, L. E. (2017). Reliable and Clinically Significant Change in Outcomes for Forensic Mental Health Inpatients: Use of the HoNOS-Secure. *International Journal of Forensic Mental Health, 16*(2), 161-171. https://doi:10.1080/14999013.2017.1286415

Dickens, G., Sugarman, P., & Walker, L. (2007). HoNOS-secure: A reliable outcome measure for users of secure and forensic mental health services. *The Journal of Forensic Psychiatry & Psychology, 18*(4), 507-514. https://doi:10.1080/14789940701492279

Dickens, G., Sugarman, P., Picchioni, M., & Long, C. (2010). HoNOS-Secure: tracking risk and recovery for men in secure care. *The British Journal of Forensic Practice, 12*(4), 36-46. https://doi:10.5042/bjfp.2010.0613

Ellwood, P. M. (1988). Shattuck lecture--outcomes management. A technology of patient experience. *N Engl J Med, 318*(23), 1549-1556. https://doi:10.1056/nejm198806093182327

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 of the CORE-OM. *The British Journal of Psychiatry, 180*, 51-60. https://doi:10.1192/bjp.180.1.51

Finch, B., Gilligan, D. G., Halpin, S. A., & Valentine, M. E. (2017). The Short- to Medium-Term Predictive Validity of Static and Dynamic Risk-of-Violence Measures in

Medium- to Low-Secure Forensic and Civil Inpatients. *Psychiatry, Psychology & Law, 24*(3), 410-427. https://doi:10.1080/13218719.2016.1247640

Gilling McIntosh, L. (2020). An Examination of the Forensic Matrix guide to delivering psychological therapies in forensic mental health services in Scotland (Unpublished doctoral thesis). University of Edinburgh. Edinburgh, UK.

Leach, C., Lucock, M., Barkham, M., Noble, R., Clarke, L., & Iveson, S. (2005). Assessing risk and emotional disturbance using the CORE–OM and HoNOS outcome measures at the interface between primary and secondary mental healthcare. *Psychiatric Bulletin, 29*(11), 419-422. https://doi:10.1192/pb.29.11.419

Macdonald, A. J., & Fugard, A. J. (2015). Routine mental health outcome measurement in the UK. *International Review of Psychiatry, 27*(4), 306-319. https://doi:10.3109/09540261.2015.1015505

McCloskey, H. (2001). Evaluation of the CORE outcome measure in a therapeutic forensic setting. *The British Journal of Forensic Practice, 3*(2), 22-28. https://doi:10.1108/14636646200100011

McGrath, K., Shaw, J., & Farquharson, L. (2020). Barriers to accessing psychological treatment for medium to high risk male young offenders. *The Journal of Forensic Psychiatry & Psychology* 31, (4), 596-612. https://doi.org/10.1080/14789949.2020.1782969

O'Connor, G., & Morris, R. (2019). The CORE-10 in screening for current mental health problems and severe mental illness in prisoners. *Criminal Behaviour and Ment Health, 29*(1), 43-46. https://doi:10.1002/cbm.2101

Ogloff, J. R., Talevski, D., Lemphers, A., Wood, M., & Simmons, M. (2015). Co-occurring mental illness, substance use disorders, and antisocial personality disorder among

clients of forensic mental health services. *Psychiatric Rehabilitation Journal, 38*(1), 16-23. https://doi:10.1037/prj0000088

Perry, T. (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.* (Doctoral Thesis). University of Sheffield. Sheffield, United Kingdom.

Perry, T., Barkham, M., & Evans, C. (2013). The CORE-OM and CORE-OM (SV) in secure settings: A template analysis of the experiences of male patients and their staff. *The Journal of Forensic Practice*, *15*(1), 32-43. https://doi.org/10.1108/14636641311299068

Pirkis, J., Burgess, P., Coombs, T., Clarke, A., Jones-Ellis, D., & Dickson, R. (2005). Routine measurement of outcomes in Australia's public sector mental health services. *Australia and New Zealand Health Policy, 2*(1), 8. https://doi:10.1186/1743-8462-2-8

Shinkfield, G., & Ogloff, J. (2014). A Review and Analysis of Routine Outcome Measures for Forensic Mental Health Services. *International Journal of Forensic Mental Health, 13*(3), 252-271. https://doi:10.1080/14999013.2014.939788

Shinkfield, G., & Ogloff, J. (2016). Comparison of HoNOS and HoNOS-Secure in a forensic mental health hospital. *The Journal of Forensic Psychiatry & Psychology, 27*(6), 867-885. https://doi:10.1080/14789949.2016.1244278

Stein, G. S. (1999). Usefulness of the Health of the Nation Outcome Scales. *The British Journal of Psychiatry, 174*(5), 375-377. https://doi:10.1192/bjp.174.5.375

Sugarman, P., Walker, L., & Dickens, G. (2009). Managing outcome performance in mental health using HoNOS: experience at St Andrew's Healthcare. *Psychiatric Bulletin, 33*(8), 285-288. https://doi:10.1192/pb.bp.108.022475

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

Table 1

Descriptive statistics of HoNOS-Secure, CORE-OM, PREOM

Scale Name	First Date of Assessment mean	First Date of Assessment Sd	Last Date of Assessment Mean	Last Date of Assessment Sd	Cronbach's Alpha
HoNOS-Secure (*n=865)*					
HONOS Total Score A to G	14.76	5.18	13.11	5.22	0.82
HONOS Total Score 1 to 12	10.72	6.97	9.95	6.19	0.77
CORE-OM (*n=50)*					
Wellbeing Subscale	5.75	4.61	3.59	3.22	0.78
Problems/Symptoms Subscale	15.66	12.27	8.38	7.93	0.92
Functioning Subscale	15.29	9.93	10.06	6.02	0.85
Risk Subscale	2.52	4.76	1.00	2.03	0.85
Total	39.64	29.06	22.94	17.85	0.95
PREOM (*n=109)*					
Total	45.84	9.88	45.27	8.54	0.89

Source: Authors own work

Table 2

Effect of therapy on HoNOS-Secure, CORE-OM and PREOM scores

HoNOS-Secure				
	Total A to G		**Total 1 to 12**	
Effect	*Estimates (S.E)*	*p*	*Estimates (S.E)*	*p*
Intercept	14.76 (0.16)	<0.001	10.16 (0.19)	<0.001
Assessed after start of therapy	-0.72 (0.07)	<0.001	-0.37 (0.11)	<0.001

CORE-OM										
	Wellbeing		**Problems/Symptoms**		**Functioning**		**Risk**		**Total**	
Effect	*Estimates (S.E)*	*p*	*Estimates (S.E)*	*p*	*Estimates (S.E)*	*p*	*Estimates (S.E)*	*p*	*Estimates (S.E)*	*p*
Intercept	5.03 (0.56)	<0.001	14.25 (1.53)	<0.001	13.49 (1.14)	<0.001	1.92 (0.59)	0.001	35.11 (3.5)	<0.001
Assessed after start of therapy	-0.54 (0.31)	0.086	-2.04 (1.24)	0.101	-1.34 (0.69)	0.052	-0.73 (0.41)	0.078	-4.66 (2.6)	0.073

PREOM		
	Total Score	
Effect	*Estimates (S.E)*	*p*
Intercept	45.16 (0.87)	<0.001
Assessed after start of therapy	0.45 (0.29)	0.122

Source: Authors own work

Table 3

Change in HoNOS-Secure and PREOM scores over admission

HoNOS-Secure				
	Total A to G		**Total 1 to 12**	
Effect	*Estimates (S.E)*	*p*	*Estimates (S.E)*	*p*
Intercept	14.77 (0.16)	<0.001	14.77 (0.16)	<0.001
Most recent assessment	-1.58 (0.15)	<0.001	-0.37 (0.11)	<0.001

PREOM		
	Total Score	
Effect	*Estimates (S.E)*	*p*
Intercept	45.25 (0.97)	<0.001
Most recent assessment	0.74 (1.02)	0.122

Source: Authors own work