A critical analysis of clinical evidence from high secure forensic inpatient services

James M. Tapp¹,², MSc.; Fiona Warren¹, PhD.; Chris Fife-Schaw¹, PhD.; Derek Perkins¹,², PhD.; Estelle Moore², PhD.

¹ Psychology Department, University of Surrey, Guildford, Surrey, GU2 7XH.

² Psychological Services, Broadmoor Hospital, Crowthorne, Berkshire, RG45 7EG.

The authors report no financial affiliations or other relationships relevant to the subject of this article.
The authors report no financial affiliations or other relationships relevant to the subject of this article.

We thank Dr Robert Morgan (RM), Dr Mike Ferriter, Dr Nick Huband and Professor Pamela Taylor (PT) for advice in the protocol formulation stages. We also thank the MSc Forensic Psychology course leaders along with high secure service professionals who assisted in identifying fugitive literature and to Dr Linda Morison and Dr Gwen Adshead for comments on a completed version of the review.
Abstract

High secure forensic inpatient hospitals serve important functions in the detention and rehabilitation of people with serious mental health needs who present as a risk to society. Establishing whether services are effective in restoring mental health and reducing risk is an important task, but one that has not been systematically undertaken. A systematic review of outcome evidence from this setting was therefore conducted to establish its scope and quality. Evidence was found from twenty-two studies for nine different interventions targeting multiple and overlapping outcomes, which included recidivism, mental health, aggression, social functioning and quality of life. Studies were commonly assessed as being at a potentially high risk of bias from validity threats. The ‘best’ available evidence was for medication, psychoeducation and third-wave cognitive-behavioural interventions. Clinical implications are tentatively offered given the narrow focus of the review on high security. Directions for the high secure services research agenda are more definitively proposed.
Introduction

High-secure forensic inpatient services are commonplace among international mental health systems (Ogloff, Roesch, & Eaves, 2000) providing specialist care for people with enduring mental health problems who concurrently pose a significant risk to society and in some instances themselves. These services have more than one primary task. They have responsibilities to safeguard the individual whilst resident in high security; and for the safety of the public; both achieved through physical and procedural security. They also have dual rehabilitative roles: to provide interventions to restore mental well-being and to reduce the risk posed by individuals in preparation for their discharge to conditions of lower security. The interventions, systems and procedures involved in delivering high secure forensic inpatient care to meet these multiple roles are numerous and multi-modal (Blackburn, 2004); the need to establish reliable outcomes is therefore also essential.

Determining whether high secure services are effective is imperative to supporting the principle of evidence-based practice for forensic psychiatric populations (Cohen & Eastman, 2000), and to promote the allocation of resources where the ability to benefit from interventions has been established (NHS Management Executive, 1991). The complex clinical and functional status of patients, and the range of interventions in high security, dictates that rehabilitation is a process of individual and collective therapeutic effort (Lindqvist & Skipworth, 2000). The focus of evaluations
of effectiveness will therefore likely be wide ranging across and between interventions and indicators of effectiveness. The need for diversion of offenders with mental health needs to appropriate resources is indicated by a range of stakeholders; including the public and service users themselves (Bradley, 2009).

Overviews of ‘what works’ in the forensic mental health arena, which by implication extends to high security, have been conducted with the objective of establishing the scope and quality of effectiveness research for secure service provision (Blackburn, 2004; Hodgins, 2002; Knabb, Welsh, & Graham-Howard, 2011; Quinsey, 1988; Rice & Harris, 1997). These have the potential to provide forensic mental health practitioners with evidence that practices with this complex patient group produce positive outcomes, and also establish research priorities for improving this knowledge base. Consistently, overviews have reported the absence of outcome research, or where it is available, that it lacks methodological rigor (Quinsey, Cyr, & Lavallee, 1988; Rice & Harris, 1997), which restricts generalizations about rehabilitation efficacy (Hodgins, 2002). Consequently, calls for new and improved research have been echoed across overviews.

Focusing on high secure services specifically, it might therefore be concluded that there is limited evidence to indicate ‘what works’ for this forensic inpatient population. However, this conclusion might be considered with caution for a number of reasons. Primarily, it is not apparent that
systematic approaches to collating forensic mental health research evidence to date have been applied, and it cannot therefore be claimed findings to date are based on all the available outcome research. A significant fugitive literature base in the forensic mental health arena (Fazel, Grann, & Långström, 2009) emphasizes this issue, particularly in reviews where this was excluded (Knabb et al., 2011).

The broad focus on forensic mental health research also limits the extent to which conclusions can be explicitly made for specific levels of service provision. Whilst theoretically it could be argued that it should be possible to extrapolate clinical evidence from different levels of secure and non-secure inpatient provision, given that high secure inpatients are admitted from and discharged to such services, marked differences exist between settings for obvious reasons, such as the therapeutic uses of security (environmental, procedural, relational), provision, timing and intensity of interventions, and related legal governance systems (Lindqvist et al., 2009). Demarcating these difference is further complicated by the absence of agreed definitions of differing levels of secure care (Crichton, 2009). Epidemiological differences support the strategy of characterizing high secure hospital residents as different from other populations: suffering from more enduring mental disorders and as presenting a higher risk of harm to others or themselves than those within less secure hospitals or prisons (Berry et al., 2003; Marshall & Willmot, 2011; Pimm, Stewart, Lawrie, & Thomson, 2004; Smith, White, & MacCall, 2004).
The proposition that outcome evidence from interventions with offender populations might also inform approaches for rehabilitating forensic patients is equally uncertain. Supporters of this position have acknowledged that there has been limited empirical support to confirm the use of information in this way (Rice & Harris, 1997). This is a status that has remained unchanged in more recent reviews on the use of offending behaviour programmes with individuals with mental disorders, which have recommended more robust research is needed to be confident as to the value of these intervention within secure mental health services (Nagi & Davies, 2010; Rees-Jones, 2011; Sainsbury Centre, 2008). This is before considering any ideological differences in relational security and rehabilitation between prison and hospital services (Gunn, 2000), which may impact on outcomes.

In view of the implications for understanding ‘what works’ for high secure forensic inpatients and building on gaps in the existing research literature, a focused and systematic review of outcome evidence from high security was indicated. Previous approaches to systematically reviewing the forensic mental health literature have synthesized research evidence for forensic patients on, epidemiology (Badger, Nursten, Williams, & Woodward, 1999), service provision (Hagell & Bourke Dowling, 1999; Lart, Payne, Beaumont, MacDonald, & Mistry, 1999) and group and therapeutic community interventions (Duncan, Nicol, Alastair, & Dalgleish, 2006; Lees, Manning,
& Rawlings, 1999). There has been no systematic review focused on the interventions provided in high secure forensic inpatient hospitals. This paper reports a critical review of outcome studies conducted in high secure hospital services in order to identify and evaluate the current evidence base. The aim is to provide forensic clinicians with an up to date and comprehensive review of clinical evidence generated from research in this setting.
Method

Practice and method guidelines for conducting a systematic evaluation of the effects of health interventions were followed (Centre for Reviews and Dissemination, 2009; Egger, Davey, & Altman, 2001; Petticrew & Roberts, 2006). The criteria for study selection were defined by Population, Intervention, Comparison and Outcome (PICO) (Booth & Fry-Smith, 2003) parameters to formulate the research question clearly and facilitate the search strategy by identifying key concepts (Villanueva, Burrows, Fennessy, Rajendran, & Anderson, 2001).

Literature search

The electronic databases accessed for retrieving studies included: AMED; ASSIA; BNI; CINAHL; DIPEX; EMBASE; HMIC; Medline; PsycINFO; SCI-E; and COPAC. Dedicated fugitive literature (research that exists outside of peer-reviewed journals and books) databases were searched (National Research Register; PsycEXTRA) to reduce publication bias. The reference lists of all included studies were searched to identify any missing research. Clinical and research practitioners within the UK and international forensic mental health arena, identified from discussion within the review team, were contacted to request relevant research evidence. High secure forensic inpatient service professional groups (e.g. Occupational Therapy; Psychological Services; Social Services; Medical Services) and Research & Development departments and Forensic Psychology MSc course co-ordinators within the UK were surveyed to further reduce the
risk of publication bias. To capture fugitive literature from international high secure forensic inpatient services a list of hospitals was initially identified from: experts in the field (PT; RM); literature reviews and searches on international forensic mental health systems (Ogloff et al., 2000); inquiries to the World Health Organisation (WHO), American Psychiatric Association (APA) and the International Association of Forensic Mental Health Services. No comprehensive source of information to identify international high or maximum secure hospital services was identified, however a number of equivalent international (Africa, America, Australia, Brazil, Canada, China, Denmark, Finland, Germany, Netherlands, New Zealand, Norway, Poland, Sweden) services were identified from experts and literature searches which were included in search terms (a list is accessible from the corresponding author). Where provided and accessible international service websites were accessed and fugitive literature requests were made. Requests were also made to international experts. No language restrictions were imposed in the selection criteria to include international studies. Database alerts (ZETOC) were created to retrieve new studies published during the review process.

No publication period was imposed as no existing reviews investigating outcomes for high secure settings were identified from systematic review databases: DARE; Campbell Collaboration; DoPHER. Dates ranged from inception of respective databases up to June 2010. Non-English studies were included within the search strategy to maximize the sensitivity of the review.
Search terms representing the research question were selected using the authors’ knowledge; existing literature; indexes and thesauruses of the literature search database systems and in consultation with research-active contributors to the topic area. Terms were mapped onto established Medical Subject Headings (MeSH) to ensure sensitivity. The full search strategy is available from the corresponding author.

**Study Selection**

Studies were eligible for inclusion in the review if they met the following Inclusion and exclusion criteria: (Medcalc Software Version 12.2.1) Randomized controlled trials (RCTs) and non-randomized studies, which reported a comparison or control group (e.g. quasi-experimental design), were included. Outcome studies with no reported comparison (or control) group were excluded; (2) Studies of adult (>18 years) detainees in high secure forensic inpatient services internationally were included. Whilst service provision is governed under different mental health legislation in different countries, consistent practices may exist between them (Ogloff et al., 2000) and needs and interventions are less likely to ‘know national borders’ (Lindqvist & Skipworth, 2000). Studies of forensic inpatients detained in lower-secure psychiatric services or of offenders with mental health needs held in correctional systems were not included because of the differences in the secure environment and the nature of intervention. Studies reporting on participants under the age of 18 years or detained in juvenile services were excluded; (Barkham
et al.) Studies of the effectiveness of any intervention were included except studies focused only on intervening with iatrogenic factors; (4) Reports describing interventions but not reporting outcomes were excluded. Outcome was defined as ‘an aspect of a participant’s clinical or functional status that changes as a result of an intervention’ (Centre for Reviews and Dissemination, 2009; p, 269). Primary (e.g. affective, cognitive and behavioral) and secondary (e.g. recidivism, re-admission and mortality) outcomes were included in the review to incorporate the widest range of potential intervention targets.

Exclusions were applied in two stages. Firstly, titles, abstracts and keywords were checked against the inclusion and exclusion criteria to assess suitability. References with insufficient criteria information were retained to reduce the risk of excluding relevant studies. The consistency of screening at this stage was checked on a randomly selected number of studies (n=30; 3%) by a rater independent of the research team and blind to this exclusion stage. The level of agreement was ‘good’ (Altman, 1991) at 79%. Secondly, remaining papers were obtained in full and the paper content checked against the inclusion and exclusion criteria. Where insufficient information was available to apply inclusion and exclusion criteria, authors were contacted.
Data extraction

Narrative synthesis was conducted to provide a summary of participant, intervention, comparison, and outcomes and facilitate the assessment of evidence (Popay et al., 2006) between reported results for interventions. A detailed narrative synthesis is available from the corresponding author.

Intervention effect sizes estimates (ESE) were calculated, where the data permitted, alongside the study narratives. Standardized mean difference (SMD) was calculated for continuous outcomes using Cohen’s d (Cohen, 1977) conventions (small effect = 0.2; medium effect = 0.5; large effect = 0.8) given their prominence among published studies (Ferguson, 2009). Relative risk (RR) probabilities were calculated for binary outcomes (e.g. recidivism) to demonstrate the degree of risk reduction following intervention when compared with no intervention. Precision of estimates was determined using a 95% confidence interval.

The assessment of study quality was conducted across nine validity threats (Table 1) that were derived from systematic review guidance and research methods theory to evaluate risks of bias (Begg et al., 1996; Centre for Reviews and Dissemination, 2009; Des Jarlais, Lyles, & Crepaz, 2004; Egger et al., 2001; Higgins & Green, 2008; Petticrew & Roberts, 2006; Yang & Miller, 2008). Absence of information (in the full paper or from correspondence with authors where this was not available) about validity threats was taken as absence of measures to address validity threats.
Extracted methodological information corresponding to the procedures for each respective validity threat was allocated a single point score where the data represents the absence (i.e. absence is explicitly reported; insufficient methodological information is reported) of a procedure to limit or remove the validity threat. Each procedure was recorded as present or absent to limit subjectivity in coding.

A total score was then given for each validity threat, by summing the respective procedure scores, which was then averaged out (by the number of procedures for each threat) to permit comparisons. Averages were categorized according to an ordinal ranking of low (0), medium (Medcalc Software Version 12.2.1) or high (2) risk for each of the validity threats. Finally a total quality rating was allocated to each study including all threats, ranging from 0-18, with a score of 0 indicating low risk and 18 indicating high risk across all nine validity threat categories. A 'good' (Altman, 1991) level of inter-rater agreement, 64%, was reached for methodological data-extraction on a randomly-selected proportion (n=3; 14%) of included studies. Based on reported limitations on the reliability and the weighting of summary scores for quality ratings (Jüni, Witschi, Bloch, & Egger, 1999), descriptive statistics were only presented to demonstrate the variability of
study quality. A bivariate correlation (overall quality rating and publication year) was used to investigate whether research quality has improved over time. Data were screened for parametric assumptions and where violated, non-parametric tests are reported.

**Meta-analysis**

Since the scope of the review was broad, explicit and stringent study criteria were applied to support a meta-analysis of effect sizes of results where any subgroups of interventions were included. A method of stratification was applied to examine two types of heterogeneity described within systematic review guidance (Centre for Reviews and Dissemination, 2009; Higgins & Green, 2008), prior to grouping studies for effect size summary. Profiles of clinical (participants, intervention, outcome) and methodological (design, selection, quality rating) diversity determined from extracted study information were created to explore the homogeneity of studies. There was evidence to indicate that intervention effects differed substantially between studies on the basis of clinical and methodological inconsistencies, across participants, comparators, methods, outcomes and quality rating. Therefore to avoid pooled effects bias (Egger et al., 2001), and the risk of drawing incorrect conclusions (Egger & Smith, 1995; Greenhalgh, 1997), no comparisons were conducted.

**Results**
A total of 1120 studies were identified from electronic databases and 133 from fugitive literature searches (Figure 1). For a proportion of studies, most notably at the third and fourth exclusion stages, it was not possible to access sufficient information from authors to determine whether inclusion criteria were met and these studies were therefore excluded. A final set of 22 studies met the inclusion criteria.

**Interventions**

Pharmacological interventions, predominantly antipsychotic treatments, were the most prevalent (n=7). Four studies evaluated psychotherapeutic interventions, two Cognitive Behavioural Therapy (CBT), and two reporting on ‘third wave’ cognitive behavioural therapies, namely Dialectical Behaviour Therapy (DBT) and Schema Modal therapy (SMT). Three studies reported outcomes of behavioural interventions (e.g. social skills training). Three were evaluations of a therapeutic milieu, two examining the assessment process of the Dangerous and Severe Personality Disorder (DSPD) services, one investigating a therapeutic community. Single studies were found evaluating psycho-educational group work, structural changes to the physical hospital environment, cranial stimulation and a dietary program. A single study also investigated outcomes following discharge describing high security treatment as a whole as the intervention.
Outcomes

Across the 22 studies, 65 different measures of outcome were described in 33 domains, which were categorized. Categorization was informed by an existing review of forensic mental health research outcome measures (Chambers et al., 2009), and through discussion within the review team to provide a refined description of outcomes. Full details of outcome categorization are available in Table 2. In over a third of studies the outcome assessment method (e.g. self-report; observer-rated) was not reported. Where described, outcomes were predominantly measured using either self-report (n=12; 18.5%) or observer ratings (n=11; 16.9%), four studies (18.1%) used case-file reviews and three (13.6%) used interview methods.

Insert Table 2 here
Methodological quality

The average total quality assessment rating shows all studies as at high risk of bias ($\overline{x} = 13.04$; S.D. 1.78; range 0-18). Studies with the lowest risks of bias (low ratings) were those evaluating pharmacological ($\overline{x} = 12.28$), psychotherapeutic ($\overline{x} = 12.25$) and psychoeducational ($\overline{x} = 10.00$) interventions. A Spearman’s rho correlation was conducted to estimate the relationship between total quality rating and year of publication. A significant negative correlation ($r = -.561$, $p = .007$) indicated that more up to date research was of improved methodological quality (indicated by a lower rating).

In studies that employed observer-rated outcomes fewer than half provided information about the training of raters or the use of inter-rater agreement and so were rated as being at risk of testing effects. No information was given in studies using self-reported outcomes to estimate the potential for socially desirable responding. In only three (13.6%) studies was a reliability coefficient of measures reported and only two (9.0%) studies reported validity data. A high degree of methodological heterogeneity was noted where both within and between interventions a range of different outcomes measures were employed to assess change in clinical and functional status. Details of the 22 studies across PICO descriptors, ESE and quality ratings are reported in Table 3.
A further breakdown of the validity threats for all studies (Table 4) demonstrates that the majority (n=18; 81%) were rated as high risk across seven of the nine validity threats. Studies were rated as being at low risk of maturation effects if there was a control group, which was one of the criteria for inclusion. An absence of information about parallel interventions exposed to outside of the target intervention, by either control or experimental groups, resulted in a medium risk level rating for history effects. Procedures to reduce selection bias of participants to conditions were reported in 12 (55%) of the studies, with the remaining studies being categorized at a high potential risk of biased recruitment and allocation. The highest possible risks were reported for mortality threats; with no studies reporting the presence or absence of missing data, drop-out comparisons or external validity.

Study findings

High Secure Hospital Treatment

In the only study investigating high secure hospital treatment holistically, a risk reduction in general recidivism (RR = 0.74) was reported for those receiving treatment in high security (n =
214) when compared to a matched comparison group assessed in high security but not admitted for treatment (n = 215) (Rice, Harris, Lang, & Bell, 1990). A similar risk reduction (RR = 0.68) was reported for violent recidivism where those at risk in the hospital treatment group (n = 209) were significantly less likely to recidivate than comparison subjects (n = 214). As no indication was given as to what specific interventions were received during admission the finding indicates the potential value of high secure hospital treatment, but determining how it achieves this could not be determined with any clarity. Also no discussion was offered on the potential limitations of investigating recidivism outcomes (Tarling, 1993).

**Milieu Interventions**

Therapeutic processes of an assessment procedure in specialist services for high secure dangerous and severe personality disordered (DSPD) inpatients (n = 21) indicated modest SMD changes (<0.5) on outcomes for social functioning, quality of life and aggression (Barrett et al., 2009), when compared with a prison referral waiting list group (n = 19). Within the same setting, using a non-conventional early (n = 33) versus late (n = 38) admission comparison group no significant differences were found (Tyrer et al., 2009). Both studies reported, but did not describe, compromised randomization processes, increasing the potential risk of selection bias. As only a brief description of the assessment process was provided it was also difficult to interpret changes in reported clinical and functional status. Recidivism outcomes following TC residency (n = 169)
(Rice, Harris, & Cormier, 1992) when compared with a non-TC group (n = 136) demonstrated relative risk reductions for violent (RR = 0.84) and general (RR = 0.83) re-offence rates. However, for TC participants assessed as having psychopathic characteristics a higher rate of recidivism was reported, indicating a potential ineffectiveness for this specific population. Non-TC comparisons were closely matched to minimize selection bias. The limitations of attributing distal outcomes to the TC model were not considered.
Investigating aspects of the high secure environment

Contrary to prediction, a procedural security intervention in the form of CCTV use on hospital wards, was associated with no significant detriment in self-reported mental health symptoms when compared to non-CCTV monitored wards (Vartiainen & Hakola, 1994). Improvements were also noted in ward atmosphere perceptions and a reduction in incidents was reported when compared with control wards without CCTV. However, no pre-intervention comparisons were conducted between wards to determine baseline ratings or differences and the use of an unstandardized assessment tool compromises the validity of the measurement of self-report symptoms.

Behavioural interventions

Mixed evidence has been reported for the effectiveness of behavioural interventions. Social skills training was associated with significant improvements in managing role-played challenging situations over non-directive psychotherapy for a group of male fire setters (n = 10) (Rice & Chaplin, 1979). In a similar evaluation of social skills training for patients (n = 16) in the same hospital, which included the use of Contingency Management (token reward system), compared against Non-directive psychotherapy (n = 18), significant improvements were also found in interpersonal functioning, however, these changes were not observed to have been maintained after the programme (Rice, 1983). Insufficient information was available to determine effect sizes.
for changes in behaviour. Blind assignment to conditions limited the risk of selection bias and
observer-ratings of social interaction of instrumentation effects at baseline. However, experimenters were aware of participant allocation following baseline assessment and care staff provided token reinforcement without control for potential influences of the distribution of these. The cross-over design reduced potential risks of history and maturation effects. Aversion therapy, applying classical conditioning, bio-feedback and signaled punishment, for extra-familial child molesters \( (n = 29) \) with a personality disorder, had no significant impact on sexual and general recidivism when compared with a control group \( (n = 29) \) (Rice, Quinsey, & Harris, 1991). The limitations of measuring recidivism or the confounding factors that may influence this outcome were not discussed.

**Psycho-educational Interventions**

Evidence for the benefits from psycho-education with inpatients with a primary diagnosis of schizophrenia \( (n = 7) \) is provided from a single study including a Treatment-as-Usual control group \( (n = 8) \) (Aho-Mustonen, Miettinen, Koivisto, Timonen, & Räty, 2008). The most prominent intervention effects were reported for improvements in illness knowledge \( (SMD = 0.54) \), awareness of illness \( (SMD = 0.30) \) and attitudes towards treatment \( (SMD = 0.27) \). Participants were not randomly allocated to conditions, increasing risk of selection bias, but were matched on baseline assessments. The study was preliminary and therefore had a small sample and low
power. Intervention integrity was supported by manual adherence and practice guidelines.

Caution concerning instrumentation effects is warranted as assessments were unstandardized and adapted by investigators.

**Psychological interventions**

Psychotherapy interventions broadly targeted offence-focused needs or mental health restoration.

Studies of CBT for adult and child sexual offenders (Marques, Day, Nelson, & West, 1994; Nicholaichuk, Gordon, Gu, & Wong, 2000) reported moderate risk reductions in recidivism rates (violent and sexual), and indicated that attrition is associated with poorest outcomes with a relative risk reduction of 0.14 for both violent and sexual offenders that completed groupwork \((n = 98)\) over those who did not \((n = 8)\) (Marques et al., 1994). Authors acknowledged potential history threats from parallel non-groupwork interventions. The integrity of the intervention studied was not readily assessable and the risk of over / under reporting of recidivism was not considered.

A study of Dialectical Behaviour Therapy \((n = 8)\) was compared with TAU \((n = 9)\) in a high security setting (Evershed et al., 2003) and significant changes in self-reported aggression and observed institutional behaviours were observed both mid and post intervention. Most marked changes occurred in violent behaviours \((SMD = 1.39)\) and violence seriousness \((SMD = 3.29)\)
post group (18 months). Potential risks of history effects and sampling bias were acknowledged from the TAU group recruitment occurring 12 months following completion of the DBT and groups not being equivalent at intake. Intervention integrity was supported through supervision of manual adherence.

A study of Schema Modal Therapy (SMT) for high secure personality disordered inpatients \((n = 29)\) reports significant SMD reductions in violence risk \((SMD = 3.43)\) and self-reported maladaptive defectiveness and shame schema \((SMD = 2.47)\) (Tarrier et al., 2010). However, differences were not statistically significant when compared with a TAU group \((n = 34)\). Allocation procedures were clearly described, reducing the risk of selection bias. A number of participants withdrew prior to randomization potentially increasing mortality risk. Instrumentation effects were controlled for with blind ratings of outcomes. Efforts were made to promote intervention fidelity including, supervision and assessment of therapist contributions, the use of a treatment protocol and expert advice. A high number of outcome assessments were applied and not directly corrected for, increasing the potential risk of a Type I error. However authors provided adjustment estimates for consideration.

**Pharmacological interventions**
Evidence for pharmacological interventions is predominantly reported for the use of antipsychotic medications for individuals with treatment resistant schizophrenia. Two studies (Dalal, Larkin, Leese, & Taylor, 1999; Swinton & Haddock, 2000) investigating the efficacy of Clozapine compared to a non-Clozapine group reported modest improvements in discharge rates at two years respectively (SMD = 0.58; RR = 2.45). A further follow up at 4 years was reported for the former study (RR = 1.96) in addition to positive reductions in violent behaviour (RR = 1.24). In this study the lack of randomization increases possible bias in recruitment for treatment (n = 50) and control (n = 50) conditions. Instrumentation effects were controlled for through inter-rater reliability. In the latter study, controls (n = 106) were matched on demographics to the treatment group (n = 106) at the time of admission to reduce history or maturation effects, and to ensure integrity of the control group. Control group membership assignment was random, but no description of the randomization process was provided to evaluate this and the study was retrospective. Differences in MHA classification were also noted by the authors, which may influence discharge decisions. Evidence of combined Lamotrigine (antiepileptic) and Clozapine medication (Tiihonen et al., 2003), compared with a placebo in a randomized cross-over trial, reported moderate SMD changes on psychiatric symptoms (SMD = 0.50), and more specifically positive symptoms (SMD = 0.28) and general psychopathology (SMD = 0.28). Participants were matched on clinical and demographic characteristics prior to the trial and a "wash out" period was incorporated to ensure no medication spillover effects. To limit instrumentation effects outcomes
were inter-rated by observers for a proportion of the participants however no description of the psychometric properties of the outcome measures is provided.

Outcomes comparing patient groups on high (n = 32) versus low (n = 32) doses of Chlorpromazine (Tavernor, Swinton, & Tavernor, 2000), reported most marked and negative SMD differences in the high dose group on thinking disturbance (SMD = 0.95) irritability (SMD = 0.88) and reported incidents of aggression over the past month (SMD = 0.63). High and low dosage groups were matched on demographics; however, as symptom and function baseline characteristics were not recorded, concluding outcomes on the basis of high versus low dosage is uncertain. To limit instrumentation effects observer ratings were made blind on one outcome variable, no information was given for the remaining observer-rated outcomes. Additional medication being taken by participants as well as the experimental antipsychotics was reported and there was no significant difference between controls and cases in terms of distribution of additional medication use, therefore, risk of selection bias was minimized. A single study investigating changes in social functioning following Risperidone (n = 10) (Beck et al., 1997) versus traditional neuroleptics (n = 10) showed no significant differences between medication types. Insufficient data were presented to determine ESE.
A retrospective case control mirror image study investigated the benefits of Topiramate antiepileptic medication, compared with Valproate antiepileptic medication and a combination of both, for patients exhibiting aggressive behavior (Gobbi, Gaudreau, & Leblanc, 2006). Topiramate was equally effective at reducing incidents of physical aggression towards others at eight weeks. At 12 weeks Topiramate was also comparably effective in reducing incidents of overt aggression, with significant reductions over Valproate and combination medication in physical aggression towards objects. Prior to treatment no significant differences were found between the groups, except for in length of hospitalization. Observer rated aggression was inter-rated to ensure reliability, however, observations of social interactions were not. Outcomes were mostly evaluated in a blind manner to reduce instrumentation effects. Drop-outs were reported but mortality effects were not considered.

A single randomized cross-over trial investigated the effects of Benperidol, a neuroleptic medication, on sexual drive and arousal among sex offenders compared with a control drug with matched side-effects (Chlorpromazine) and a placebo (Tennent, Bancroft, & Cass, 1974). Statistically significant positive changes were reported in sexual interests following Benperidol and sexual attitudes following Chlorpromazine. However, insufficient data were presented to determine ESE. No indication was given of the risk of cross-over contamination effects and no time lapse in between administrations was reported. Information concerning parallel interventions
was also absent and potential instrumentation effects with self-reported sexual activity warrant consideration.

Neurophysiological Interventions

A single study examined the effect of cranial stimulation \((n = 11)\) compared with a sham condition \((n = 11)\) on psychotic symptoms using double-blind random-assignment (Holi et al., 2004). No significant improvements and minimal SMD changes in symptom experience were reported. The most marked improvements were reported for the sham group. Sham and stimulation groups were matched on a series of characteristics (including drug regimes) to minimize selection bias. Delivery of the stimulation was across two different procedures, which may have impacted on reported outcomes if participants could accurately determine allocation based on variance in procedures.

Dietary Interventions

A single study investigated the effects of diet on mental health symptoms (Vlissides, Venulet, & Jenner, 1986). A double blind cross-over trial was conducted to investigate the influence of gluten on positive symptoms of schizophrenia. Statistically significant improvements in observer-rated anxiety, depression, hostility and psychotic symptoms following the introduction of the gluten-free diet compared to a pre-trial condition were reported. The integrity of the blinding procedure was
clearly described. Participants were also observed to ensure adherence to the dietary changes, however no information was given about inter-rater reliability of observations of outcomes.

**Discussion**

Interventions delivered during detention in high security services are necessarily wide-ranging in target as the model of care is multi-modal given case complexity. Reported intervention effect sizes can only be interpreted with caution given the potential risk of inflated estimates associated with limited methodological quality (Baguley, 2009; Ferguson, 2009; Olejnik & Algina, 2003). However, positive outcomes are evident in some studies with improved designs that can provide useful practice and research implications.

There is evidence for pharmacological interventions, with reports of benefits from the use of Clozapine with patients who have treatment resistant schizophrenia (Dalal et al., 1999; Swinton & Haddock, 2000; Tiihonen et al., 2003); findings that are congruent with existing UK practice guidance (National Institute for Clinical Excellence, 2009). Preliminary evidence for psycho-education for those with mental illness (including schizophrenia) indicates improvements in insight and therapy engagement, emphasizing the interaction effects of interventions in the high secure setting (Aho-Mustonen et al., 2008). ‘Third wave’ cognitive-behavioural therapies offer promising evidence of risk reduction for detainees with a history of serious interpersonal violence
(Evershed et al., 2003; Tarrier et al., 2010). However, complications with determining quantifiable benefits over TAU, which represents nested interventions within which there are problems of disentangling multiple interacting therapeutic ingredients, limit confidence in comparison (Tarrier et al., 2010).

With regard to the exploration of non-findings, researchers have shared the observation that mis-timed interventions can be harmful and promote deterioration in those known to be highly vulnerable (Jones, 2007). This was demonstrated by the increased risk in recidivism from non-completion of interventions. The absence of outcome studies that target common clinical features of the high secure population (Jackson, 2000; Scott, 2004) also highlights a shortage of research evidence for intervening with specific complex needs and that this warrants attention.

Given the specific focus of the review it is important to emphasise that whilst the findings aimed to provide a comprehensive summary of evidence that describes effective clinical practices in high security, absence of evidence for interventions does not suggest these are ineffective. Equally, the exclusion of evidence from other related settings (e.g. prison, lower secure or non-forensic mental health services) does not mean that this information could not be extrapolated to high security for the purposes of informing evidence-based practice. However, it is recommended that empirical support for this practice needs to be established. What can be concluded is that the
findings collated here do indicate there is limited outcome information derived from this population and this raises a number of important research implications, which are discussed below.

**Implications for outcome research in high security**

Several recommendations can be drawn with respect to outcome research in high secure forensic inpatient services. It is imperative that the quality of outcome evaluation is improved to increase confidence in what is a growing evidence base. The absence of RCT evaluations of interventions in high security is noted and the quality control for RCTs that are conducted should be comprehensive in the management of risks of bias (Slade & Priebe, 2001). This recommendation is supported by the fact that the quality ratings of studies were higher, in some instances, for quasi-experimental and observational studies than for some RCTs, which were rated as being at greater risk from specific validity threats such as instrumentation effects, mortality and intervention fidelity. Whether this was due to incomplete reporting or methodological limitations is unknown but an acknowledgement by outcome investigators of these issues is essential to support internal validity assessment. Consistency in methods of reporting has been variable, which compromises the task for evaluators, journal editors and professionals involved in the wider dissemination of outcome research in this area, (Begg et al., 1996; Des Jarlais et al., 2004).
An additional implication for RCT methods from the review findings concerns the feasibility of implementing an RCT design in an applied forensic mental health setting, where defining and measuring the impact of nested interventions targeting a range of outcomes (e.g. biological, psychological, social), is complicated. The protocol violations reported in the trials reviewed spanned a range of procedural problems (e.g. recruitment, allocation, intervention implementation), resulting in risks of biases such as history effects, maturation, intervention fidelity and mortality. Such complications support recommendations to investigate and control barriers to conducting trials prior to their implementation (Farrington & Jolliffe, 2002); and guidance exists to aid achieving this (Pope & Mays, 1995). Equally the value of evidence from non-RCT research might be considered even where research is of limited methodological rigour but reports comparable outcomes, for example the benefits of therapy engagement following psychoeducation (Vallentine, Tapp, Dudley, Wilson, & Moore, 2010). The reliability and validity of measures of change is deemed a fundamental requirement for intervention research (Kline, 2000) yet information about this was infrequently provided in the studies. The confidence in outcome measures could only be assessed on the available evidence and is, therefore, low in the majority of studies reviewed. Standardization of outcome measures for forensic patients is recommended as a significant investment for evaluators to promote confidence in reported changes. This includes recidivism which is a complex measure of success potentially influenced by multiple confounding factors (Gunn & Taylor, 1993). Such efforts will also facilitate the summary of
findings and remove the need for multiple outcomes testing which may inflate Type I error risk, as was evident in a number of the studies reviewed. Triangulation of multiple perspectives (self-report, interviewer, observational) is also recommended to reduce testing effects through identifying discrepancies between observer and participant reported outcomes.
Research Recommendations

The need for methodologically rigorous outcome studies remains critical for high-secure forensic inpatient services. A series of research recommendations are offered on the basis of the findings from this review.

1. Adhere to standardized guidelines (Begg et al., 1996; Des Jarlais et al., 2004) when reporting evaluation findings to ensure methods are transparent. This would support the validity assessment and summation of research in an area characterized by low base-rate problem behaviours and multiple and complex validity threats.

2. Based on the existing literature, investigate the feasibility and define the barriers for RCT implementation in high-secure settings to inform future evaluations. Evidence that certain non-RCT design studies (Aho-Mustonen et al., 2008) were rated of higher quality compared to RCT designs also demonstrates the potential value in including alternative forms of evidence in informing clinical practice, including the follow up of forensic patients (Buchanan, 1998).
3. Identify history threats (e.g. institutional effects, ‘active rehabilitation ingredients’) inherent in the high security environment to inform evaluations where randomization is not applied, and record potential moderating effects on outcomes.

4. Include outcome measures with a history of application in relevant clinical settings and with established psychometric properties.

**Review limitations**

Whilst efforts were made to maximize literature search sensitivity in this review several limitations need to be considered. Despite generic search terms to identify high security, the author’s familiarity with the health care structures in the UK may have restricted a comprehensive list of equivalent international high secure services. The high volume of studies excluded due to insufficient information being available also increases the risk of publication bias, as did the limited hand searching of key journals.

The application of categorical data to assess methodological quality was intended to limit subjectivity in interpretation of procedures but examining the level of inter-rater agreement for the data extraction process (64%) there were differences in the information extracted. A pilot of the data extraction schedule used to distil study information would have further improved the level of
accuracy with which data and methods of studies were recorded. In addition, the validity threats described are not exhaustive of all risks of biases to intervention studies, for example social threats from non-independent evaluators that delivered interventions.
References


<table>
<thead>
<tr>
<th>Validity threat</th>
<th>Description</th>
<th>Procedures for managing validity threats</th>
</tr>
</thead>
<tbody>
<tr>
<td>History threat</td>
<td>Events that occur outside of the target intervention that impact on the outcome</td>
<td>Control group; report parallel / co-occurring interventions</td>
</tr>
<tr>
<td>Maturation</td>
<td>Individual changes due to internal processes operating as a function of the passage of time</td>
<td>Control group</td>
</tr>
<tr>
<td>Testing effect</td>
<td>Effects of repeat testing, risk of self report bias or impression management</td>
<td>Control group; blind participants to condition; validity checks to control for response bias</td>
</tr>
<tr>
<td>Instrumentation</td>
<td>Properties of, and changes to, the outcome measurement process</td>
<td>Reliability; validity; training of outcome raters; inter-rater assessment; blind outcome assessors to conditions</td>
</tr>
<tr>
<td>Statistical regression</td>
<td>Including outliers in the evaluation who naturally regress to the mean</td>
<td>Reported outlier cases; parametric assumption tests</td>
</tr>
<tr>
<td>Selection bias</td>
<td>Biased sample-selection, including allocation to the intervention</td>
<td>Inclusion and exclusion criteria; sampling; recruitment; assignment; comparisons between groups</td>
</tr>
<tr>
<td>Mortality threat</td>
<td>Biased estimate of change resulting from treatment drop-out or missing data</td>
<td>Drop out comparisons; missing data management</td>
</tr>
<tr>
<td>Intervention fidelity</td>
<td>Ensuring the intervention is developed with a clear rationale and implemented consistently</td>
<td>Theory; guidelines; manualisation; transparency about modification; supervision; facilitator attendance</td>
</tr>
<tr>
<td>External validity</td>
<td>Results that are generalizable to the population being studied</td>
<td>Testing of representativeness of population characteristics</td>
</tr>
</tbody>
</table>
Figure 1.

Titles and abstracts identified and screened  
\( n=1120 \)

Manually extracted duplicates  
\( n=105 \)

First Stage Exclusion  
\( n=887 \)

Full copies retrieved and assessed for eligibility  
\( n=128 \)

Second Stage Exclusion  
\( n=50 \)

Third Stage Exclusion  
\( n=65 \)

No reported outcomes / interventions  
\( n=24 \)

Setting not high / maximum security  
\( n=7 \)

Multiple secure settings  
\( n=1 \)

Correctional settings  
\( n=1 \)

Unable to obtain in full (Dissertations)  
\( n=3 \)

Uncertain on level of security  
\( n=29 \)

Fourth Stage Exclusion  
\( n=124 \)

Not Eligible  
\( n=106 \)

Unable to obtain in full  
\( n=4 \)

Study ongoing  
\( n=6 \)

No response from study authors requesting full reports  
\( n=8 \)

Studies identified from reference list  
\( n=34 \)

Fugitive Literature  
\( n=133 \)

PsycExtra  
\( n=10 \)

National Research Register  
\( n=67 \)

Call for Information  
\( n=2 \)

High Secure Services Audit Databases  
\( n=18 \)

MSc Forensic Psychology Courses  
\( n=2 \)

Studies included in the review  
\( n=22 \)
Table 2. Summary of included studies across PICO descriptions and ESE

<table>
<thead>
<tr>
<th>Author / Country</th>
<th>Total (N)¹</th>
<th>Intervention category</th>
<th>Intervention type</th>
<th>Comparison</th>
<th>Outcome(s) category</th>
<th>Effect sizes SMD / RR Range²</th>
<th>Quality assessment rating³</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rice et al., 1990 Canada</td>
<td>516</td>
<td>High Secure Hospital Treatment</td>
<td>Admission</td>
<td>Non-admitted comparison group</td>
<td>Re-offending</td>
<td>0.65 - 0.74*</td>
<td>15</td>
</tr>
<tr>
<td>Rice et al., 1992 Canada</td>
<td>176</td>
<td>Milieu</td>
<td>Therapeutic Community (TC)</td>
<td>Non-TC group</td>
<td>Re-offending</td>
<td>0.83 - 0.84*</td>
<td>14</td>
</tr>
<tr>
<td>Barrett et al., 2009 UK</td>
<td>44</td>
<td>Milieu</td>
<td>Dangerous Severe Personality Disorder Service Assessment Procedure</td>
<td>Control group</td>
<td>Aggression</td>
<td>-</td>
<td>14</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Social functioning</td>
<td>0.09</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Quality of life</td>
<td>0.03</td>
<td></td>
</tr>
<tr>
<td>Tyrer et al., 2009 UK</td>
<td>75</td>
<td>Milieu</td>
<td>Dangerous Severe Personality Disorder Service Assessment Procedure</td>
<td>Late assessment group</td>
<td>Aggression</td>
<td>-</td>
<td>15</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Social functioning</td>
<td>0.08-0.64</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Quality of life</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Vartiainen et al., 1994 Finland</td>
<td>97 ≈</td>
<td>Environmental</td>
<td>CCTV monitored wards</td>
<td>Control wards</td>
<td>Ward environment</td>
<td>0.12 – 0.42</td>
<td>15</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Mental health</td>
<td>0.07 – 0.60</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>CCTV attitudes</td>
<td>0.09 – 0.43</td>
<td></td>
</tr>
<tr>
<td>Rice et al., 1979</td>
<td>10</td>
<td>Behavioural</td>
<td>Social skills / Non-directive psychotherapy</td>
<td>Control group</td>
<td>Social functioning</td>
<td>Insufficient Data</td>
<td>14</td>
</tr>
<tr>
<td>Author / Country</td>
<td>Total (N)¹</td>
<td>Intervention category</td>
<td>Intervention type</td>
<td>Comparison</td>
<td>Outcome(s) category</td>
<td>Effect sizes SMD / RR</td>
<td>Range²</td>
</tr>
<tr>
<td>-----------------</td>
<td>-----------</td>
<td>-----------------------</td>
<td>-------------------</td>
<td>---------------------</td>
<td>---------------------</td>
<td>------------------------</td>
<td>--------</td>
</tr>
<tr>
<td>Rice 1983 Canada</td>
<td>65</td>
<td>Behavioural</td>
<td>Social skills / Non-directive psychotherapy / Contingent / Non-contingent [Group]</td>
<td>Cross over trial</td>
<td>Social functioning</td>
<td>Insufficient Data</td>
<td></td>
</tr>
<tr>
<td>Rice et al., 1991 Canada</td>
<td>153</td>
<td>Behavioural</td>
<td>Aversion therapy</td>
<td>Control group</td>
<td>Re-offending</td>
<td>1.22*</td>
<td></td>
</tr>
<tr>
<td>Marques 1994 USA</td>
<td>602</td>
<td>Psychotherapy</td>
<td>Cognitive Behavioural Therapy [Group / Individual]</td>
<td>Control / Attrition group</td>
<td>Re-offending</td>
<td>0.21-0.87*</td>
<td></td>
</tr>
<tr>
<td>Nicholaichuk et al., 2000 Canada</td>
<td>579</td>
<td>Psychotherapy</td>
<td>Cognitive Behavioural Therapy [Group]</td>
<td>Comparison group</td>
<td>Re-admission</td>
<td>0.49-1.15*</td>
<td></td>
</tr>
<tr>
<td>Evershed et al., 2003 UK</td>
<td>36</td>
<td>Psychotherapy</td>
<td>Dialectical Behaviour Therapy [Group / Individual]</td>
<td>Treatment As Usual</td>
<td>Institutional behaviour Aggression</td>
<td>0.25-3.27</td>
<td>0.26-1.27</td>
</tr>
<tr>
<td>Tarrier et al., 2010 UK</td>
<td>63</td>
<td>Psychotherapy</td>
<td>Schema Modal Therapy [Group / Individual]</td>
<td>Treatment As Usual</td>
<td>Aggression Social functioning Mental health</td>
<td>0.19-3.43</td>
<td>0.10-0.73 0.01-2.47</td>
</tr>
<tr>
<td>Author / Country</td>
<td>Total (N)</td>
<td>Intervention category</td>
<td>Intervention type</td>
<td>Comparison</td>
<td>Outcome(s) category</td>
<td>Effect sizes SMD / RR Range</td>
<td>Quality assessment rating</td>
</tr>
<tr>
<td>-----------------</td>
<td>----------</td>
<td>-----------------------</td>
<td>-------------------</td>
<td>------------</td>
<td>---------------------</td>
<td>-----------------------------</td>
<td>---------------------------</td>
</tr>
<tr>
<td>Aho-Mustonen et al., 2008 Finland</td>
<td>16</td>
<td>Psychoeducation</td>
<td>Psychoeducation Group</td>
<td>Treatment As Usual</td>
<td>Mental health awareness</td>
<td>0.18–0.54</td>
<td>10</td>
</tr>
<tr>
<td>Tennent et al., 1974 UK</td>
<td>15</td>
<td>Pharmacological</td>
<td>Antipsychotic [Benperidol; Chlorpromazine; Placebo]</td>
<td>Cross over trial</td>
<td>Sexual behaviour</td>
<td>Insufficient data</td>
<td>14</td>
</tr>
<tr>
<td>Beck et al., 1997 USA</td>
<td>20</td>
<td>Pharmacological</td>
<td>Antipsychotic [Risperidone]</td>
<td>Traditional Neuroleptics</td>
<td>Social functioning</td>
<td>Insufficient Data</td>
<td>12</td>
</tr>
<tr>
<td>Dalal et al., 1999 UK</td>
<td>56</td>
<td>Pharmacological</td>
<td>Antipsychotic [Clozapine]</td>
<td>Control / Discontinued group</td>
<td>Discharge Mental health</td>
<td>1.07 7.89* 1.23*</td>
<td>13</td>
</tr>
<tr>
<td>Tavernon et al., 2000 UK</td>
<td>64</td>
<td>Pharmacological</td>
<td>Antipsychotic [Chlorpromazine or equivalent]</td>
<td>High versus Low dose</td>
<td>Social functioning</td>
<td>0.30 - 0.39 0.25</td>
<td>13</td>
</tr>
<tr>
<td>Swinton et al., 2000</td>
<td>171</td>
<td>Pharmacological</td>
<td>Antipsychotic [Clozapine]</td>
<td>Control group</td>
<td>Length of stay</td>
<td>Insufficient Data</td>
<td>13</td>
</tr>
<tr>
<td>Author / Country</td>
<td>Total (N)</td>
<td>Intervention category</td>
<td>Intervention type</td>
<td>Comparison</td>
<td>Outcome(s) category</td>
<td>Effect sizes SMD / RR Range</td>
<td>Quality assessment rating</td>
</tr>
<tr>
<td>------------------</td>
<td>----------</td>
<td>-----------------------</td>
<td>-------------------</td>
<td>------------</td>
<td>---------------------</td>
<td>-----------------------------</td>
<td>--------------------------</td>
</tr>
<tr>
<td><strong>UK</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tiihonen et al., 2003 Finland</td>
<td>68</td>
<td>Pharmacological</td>
<td>Antipsychotic [Lamotrigine with Clozapine]</td>
<td>Placebo</td>
<td>Mental health</td>
<td>0.02-0.28</td>
<td>12</td>
</tr>
<tr>
<td>Gobbi et al., 2006 Canada</td>
<td>45</td>
<td>Pharmacological</td>
<td>Antiepileptic [Topiramate; Valproate; Combination]</td>
<td>Case control</td>
<td>Aggression Mental health</td>
<td>0.06-0.57 0.12-1.02</td>
<td>9</td>
</tr>
<tr>
<td>Holi et al., 2004 Finland</td>
<td>22</td>
<td>Physiological</td>
<td>Transcranial magnetic stimulation (rTMS)</td>
<td>Sham group</td>
<td>Mental health</td>
<td>0.02-0.07</td>
<td>13</td>
</tr>
<tr>
<td>Vlissides et al., 1986 UK</td>
<td>24</td>
<td>Dietary</td>
<td>Gluten free diet</td>
<td>Pre-trial condition / Casilan diet / Gluten drink</td>
<td>Mental health</td>
<td>0.86</td>
<td>14</td>
</tr>
</tbody>
</table>

1 Total (N) refers to the overall sample to indicate the size of the study, sample sizes for sub-group analyses are described in the narrative synthesis of individual studies.

2 RR is indicated by * otherwise the effect size estimate is a Standardised Mean Difference (SMD); a (-) indicates the outcome does not favour the intervention group over the comparator group.

3 Higher score denotes lower quality.
<table>
<thead>
<tr>
<th>Validity Threat</th>
<th>Level of Risk</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Low</td>
</tr>
<tr>
<td>History</td>
<td>0</td>
</tr>
<tr>
<td>Maturation</td>
<td>22</td>
</tr>
<tr>
<td>Testing</td>
<td>0</td>
</tr>
<tr>
<td>Instrumentation</td>
<td>1</td>
</tr>
<tr>
<td>Mean Regression</td>
<td>7</td>
</tr>
<tr>
<td>Selection Bias</td>
<td>6</td>
</tr>
<tr>
<td>Mortality</td>
<td>0</td>
</tr>
<tr>
<td>Intervention Fidelity</td>
<td>0</td>
</tr>
<tr>
<td>External Validity</td>
<td>0</td>
</tr>
</tbody>
</table>
Table 4. Categories of reported study outcomes

<table>
<thead>
<tr>
<th>Category</th>
<th>Outcomes</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Re-offending</td>
<td>Recidivism</td>
<td>4</td>
</tr>
<tr>
<td>Re-admission</td>
<td>Re-admission</td>
<td>1</td>
</tr>
<tr>
<td>Mental health</td>
<td>Psychotic Symptoms</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>Depression</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Personality</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Impulsivity</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Maladaptive schema</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Cognitive ability</td>
<td>1</td>
</tr>
<tr>
<td>Sex Offending</td>
<td>Testosterone</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Sexual Interest</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Masturbation frequency</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Erectile response</td>
<td>1</td>
</tr>
<tr>
<td>Social Function</td>
<td>Social Functioning</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Behavioural Checklist</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>Interpersonal Interactions</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Social Skills</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Communication and Social Contact</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Assertiveness</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Self-isolation</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Social Adjustment</td>
<td>1</td>
</tr>
<tr>
<td>Aggression</td>
<td>Aggression</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td>Violence</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>Hostility</td>
<td>1</td>
</tr>
<tr>
<td>Self-harm</td>
<td>Self-harm</td>
<td>1</td>
</tr>
<tr>
<td>Institutional behaviour / management</td>
<td>Institutional Incidents (observer rated)</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Nursing needs</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Medication prescriptions</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Length of stay</td>
<td>1</td>
</tr>
<tr>
<td>Iatrogenic effects</td>
<td>Side effects</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Dyskinesia</td>
<td>1</td>
</tr>
<tr>
<td>Quality of Life</td>
<td>Quality of Life</td>
<td>2</td>
</tr>
<tr>
<td>Perceptions of service support</td>
<td>Perception of supportive environment</td>
<td>1</td>
</tr>
<tr>
<td>Mental Health Awareness</td>
<td>Illness Knowledge</td>
<td>2</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td></td>
<td><strong>65</strong></td>
</tr>
</tbody>
</table>