View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by Nottingham Trent Institutional Repository (IRep)

This is the final version submitted for publication in *Therapeutic Communities* prior to the publisher's edits (1st March 2016)

Outcomes of therapeutic community treatment for personality disorder

Capone, G., Schroder, T., Clarke, S.P., & Braham, L.

Abstract

Purpose - The paper reviewed quantitative research since 1999 evaluating the effectiveness of democratic therapeutic community (DTC) treatment for individuals with personality disorders (PD) with reference to interpersonal and offending risk outcomes.

Design/methodology/approach - A systematic search resulted in the review of ten studies. All of the studies investigated DTCs treating PD in community, inpatient residential and forensic settings. Only peer-reviewed, English-language articles employing a quantitative design were included.

Findings - The majority of studies were conducted poorly and of low methodological quality, with limitations located in the representativeness of participants, limited use of control and comparison groups, follow up periods and controls for confounders. Heterogeneity remained in use of measures and limited consideration was given to the validity of interpersonal measures used. While improved interpersonal outcomes post DTC treatment were noted in forensic and residential settings, results were mixed in day and mini TC settings. Inconsistent findings in offending risk outcomes were also indicated. A study with increased methodological rigour indicated residential treatment had limited effects on interpersonal outcomes, when compared to combination treatment (residential TC and step-down treatment).

Originality/value - The study provided an evaluation of the limitations of DTC research across a range of settings and highlighted a combination of residential TC and step-down treatment may achieve superior outcomes to residential TC treatment alone in a community inpatient population. Recommendations are made for future research to contribute to the treatment of PD.

Key Words Democratic therapeutic communities, Outcome, Personality disorder, Systematic review.

Paper type *Research paper*

Introduction

Personality disorder and DTC treatment

The diagnosis of personality disorder (PD) is associated with high rates of substance misuse, disproportionate service use, social disability, crime, and mortality (Banerjee, Gibbob & Huband, 2009; Fok *et al.,* 2014). Consequently, effective treatments are important for individuals and the wider community. DTCs have been commonly implemented in the treatment of personality disorder (PD) (Rutter & Tyrer, 2003). Kennard (2004, p. 296) usefully defines DTCs as a 'living-learning situation' whereby, 'difficulties a member has experienced in relations with others outside are re-experienced and reenacted, with regular opportunities—in groups, community meetings…to examine and learn from these difficulties'. A DTC is most usefully defined as a treatment modality (i.e. integrating a range of psychological and/or pharmacological approaches) as opposed to a specific treatment method itself (Kennard, 1998).

DTCs abide by Rapoport's (1960) principles, developed via ethnographic research at the Henderson Hospital. Four core principles were identified to describe the main elements of a TC environment: Democratisation – a flattened hierarchy, with members sharing equal power in decision making processes; Communalism – curious enquiry into personal difficulties of others; Permissiveness – toleration of others' behaviour to aid development of self-awareness of maladaptive responses; Reality confrontation – individuals are confronted with interpretations of their behaviour from staff and peers within the TC (Rapoport, 1960).

DTCs have been adapted to operate successfully within a range of settings to treat PD (Kennard, 2004). Within a community setting, DTCs are implemented on a part time basis - known as mini (2 days or less per week) and day TCs (3-5 days a week with no overnight facilities) in addition to more traditional long-term residential settings (Pearce & Haigh, 2008). DTCs have also been adapted and modified for use in forensic settings, and for individuals with intellectual disabilities (ID) (Newberry, 2010; Shuker, 2010; Taylor, Crowther & Bryant, 2015). Even so, a conflict between rehabilitation and psychotherapy has continued to remain in many contemporary communities (Campling, 2001).

The many faces of a TC

TCs have previously been defined in terms of two broad categories – concept and democratic TCs (Lees, Manning & Rawlings, 2004). Concept TCs were specifically designed to treat individuals with addiction difficulties, and are differentiated from DTCs via their use of a social hierarchy, with experienced residents and staff harbouring increased authority (Vandevelde *et al.,* 2004).

Variations of DTCs have also been used to treat individuals with difficulties other than PD, such as acute and long-term psychoses (Kennard, 2004). DTC principles have been translated into hospital practices catering for this population to

incorporate a more humane approach to patient care. Community based housing projects have also been developed, to support individuals with 'treatment resistant' symptoms discharged from hospital treatment within a domestic setting, and increase involvement in clinical care (Kennard, 2004). Whilst there are similarities shared between all models, DTCs for PD specifically aim to achieve social maturation and personality change (Vandevelde *et al*., 2004).

A summarised history of treatment evaluation and some methodological limitations

Although DTCs are not included within the National Institute for Clinical Excellence (NICE) guidelines for the treatment of PD (Antisocial PD – NICE, 2010; Borderline PD - NICE, 2009), they continue to be used in community and forensic settings with promising results (Lees *et al.,* 2004; Warren *et al.,* 2003). The limited 'gold standard' evidence base (otherwise known as Randomised Controlled Trials - an experimental design involving random allocation to treatment and control conditions) for this intervention compared to other developing psychotherapy treatments for this client population, such as Cognitive Behaviour Therapy (CBT) and Dialectical Behaviour Therapy (DBT) (NICE, 2009) has prevented its inclusion within treatment recommendations (Pearce & Autrique, 2010).

While DTCs have an extensive research history of equal efficacy comparative to existing treatments (Haigh, 2002), the credibility of research findings has been weakened by a number of methodological limitations. One issue pertains to heterogeneity within client samples. This was previously highlighted within an international systematic review completed to assess the efficacy of TC treatment for people with PD and mentally disordered offenders in secure and non-secure settings. A meta-analysis of 22 controlled studies (19 of which were DTCs) from 1960-1998 identified a strong positive effect for individuals attending DTCs (Lees, Manning & Rawlings, 1999). However, DTC efficacy in the treatment of PD remained unclear due to the limited percentage of participants assessed for this disorder, preventing clear operationalization of the client group (Lees *et al.*, 1999).

More recent attempts have been made to systematically review TC research conducted beyond this time period (Veale *et al*., 2014; Magor-Blatch *et al.,* 2014). However, these reviews have succumbed to similar issues regarding sample heterogeneity, or failed to distinguish between concept TCs and DTCs, limiting the application of findings. Some authors have argued problems with sample variation may be reflective of the limitations of a diagnostic approach for complex mental health difficulties as opposed to poor methodological design (Maj, 2005).

Other difficulties noted with regards to generating 'gold standard' evidence for this treatment modality have included; absence or reduced time of follow up, attrition, participant selection and randomization, and establishing a suitable control group (Lees *et al.,* 1999; Warren *et al.,* 2003). In response to the latter issue, use of waiting list controls has been advocated due to ethical and procedural difficulties noted in

assigning individuals to a modality unsuitable for their treatment needs (Warren *et al.*, 2003). The individualised nature of treatment has also limited measurement and standardization (Pearce & Autrique, 2010). In sum, application of randomized controlled trial methodology may not adequately reflect the complex nature of a DTC or its matrix of interrelated treatment components (Haigh, 2014). Research evidence supporting TCs would be usefully considered in light of these factors.

The broad implementation of DTCs has further complicated attempts at treatment definition. For example, modifications to TC principles, such as democratization, are mandated within forensic environments to ensure the safety of patients and staff (Polden, 2010). Problems in defining DTCs have raised further questions as to treatment integrity. For example, how representative is a TC of a DTC model? It is therefore imperative DTC treatment is defined in sufficient detail to allow comparison of studies and replication.

In an attempt to counter the above issues, an accreditation process was developed by the Community of Communities to provide a quality assurance network to measure standards of good practice against the above principles amongst other agreed methods and features of DTCs (Haigh & Tucker, 2004; Kennard & Lees, 2001).

What should be measured?

A multitude of different measures have been used in existing research, preventing useful interpretation of findings (Campling, 2001). Outcomes in TC research, and for psychotherapy treatment of PD more generally have included a mixture of mental health (clinician, self-report), general distress, behavioural, service use and global outcome measures (Bateman & Fonagy, 2000; Bateman & Tyrer, 2004; Gendreau, Little & Goggin, 1996; Madan & Fowler, 2015; Magor-Blatch *et al.,* 2014). Others have highlighted the significance of identifying interpersonal change post DTC treatment (Heede *et al*., 2009; Hopwood *et al.,* 2013; Shuker & Newberry, 2010), in line with a TC's main emphasis on providing a corrective emotional experience for individuals to facilitate development of adaptive ways of relating with others.

Haigh (2013) encapsulated this relational process as the provision of 'secondary emotional development', achieved via progression through five key conditions: 'attachment (belonging), containment, communication, inclusion, and agency' (Haigh, 2013, p. 6). Consequently, interpersonal and offending risk measures may be of particular value.

Aims

Since the above methodological limitations have been raised, further outcome research in DTCs for PD has been conducted. However, their combined results have yet to be explored specifically in relation to interpersonal outcomes, such as attachment style and emotional development. The aim of the present review is to offer a critical analysis of literature evaluating interpersonal and offending risk outcomes for individuals with PDs subsequent to democratic therapeutic community treatment, and question whether outcomes are influenced by treatment setting.

Method

Search Overview

Studies were identified through database searching and a hand search of the International Journal of Therapeutic Communities. Editorials, commentaries, book chapters and previous reviews were also searched for references, but were not included in the systematic review.

Eligibility criteria

To be eligible for inclusion, studies must have included participants who were described to have a PD (via screening measure or diagnosis) in day, inpatient and secure settings. All types of PD were included, as research has indicated distinct categories to be heterogeneous in nature (Widiger, 2012).Studies examining democratic TCs for individuals with other alternative primary diagnoses (e.g. psychosis) were excluded, as these studies were not relevant to the aims of this review. Studies that assessed pre and post outcomes for individuals with a PD via the following measures were also eligible. Primary outcomes: Measures that focused on interpersonal situations (actual or mentally represented) involving a self and other (Hopwood et al., 2013) - personality disorder symptoms, cognitive schema assessments, attachment type, observed/self-reported social functioning. Secondary outcome measures: Offending risk measures (risk of violent offending and reconviction risk; Coid et al., 2007) – standardized risk measures, offending risk associated personality trait measures, incidents of physical aggression.

Interpersonal relating and offending outcomes were specifically chosen due to a prioritised aim of TC treatment focusing on interpersonal change via provision of a corrective emotional experience. This aim is reflected in Haigh (2013)'s quintessence principles. As data from psychometric and behavioural measures are presented quantitatively, only studies with a quantitative design were included.

Quantitative measures focusing solely on levels of distress, self-harm, suicide attempts or mental health symptoms were also excluded. These outcomes were explored in a recent systematic review of TCs (Magor-Blatch *et al.,* 2014) It was beyond the scope of the review to examine other outcome measures of cost effectiveness such as utilization of psychiatric services. Therefore, studies that

focused exclusively on cost efficacy outcomes of democratic TCs were excluded. Only peer-reviewed journal articles were included to safeguard the quality of studies included.

Further eligibility criteria included TCs that followed a democratic TC model (Jones, 1952) (mini, day, residential and forensic TCs) to ensure a level of fidelity to this type of TC in the delivery of the treatment modality. This inclusion criterion was used in a previous review (Lees et al., 2004). Studies that focused on different models of TC (e.g. concept TCs) were excluded from the review. The review focused on studies published from 1999 to the current date (2015), as a similar review comprehensively reviewed literature on outcomes for individuals with PD post DTC treatment up until 1999 (Lees et al., 2004). For practical reasons, only English language studies were considered.

Database selection and search

Three databases were searched: PsycINFO, EMBASE, Web of Science (WoS). These databases were used in a prior comprehensive review of TCs (Lees *et al.,* 2004). All three databases cover the time period specified (1999-2015), and include large numbers of journals, which are relevant to the review (see Table 1 below). A systematic review of medical and social science databases was undertaken. Search terms included (therapeutic community, social therapy, milieu therapy, prison therapy) and (personality disorder, outcome, efficacy, evaluation, conviction, reconviction, reoffending, recidivism).

Table 1. A description of the databases searched (as cited by: OvidSP, 2015; Web of Science, 2015).

Database	**Dates covered**	**Number of journals**	**Topics covered**
EMBASE	1974 – 2015 Week: 24	More than 7,500 (peer-reviewed journals)	Health and medical sciences including subjects of interest to

			Psychologists.
PsycINFO	1806 – 2015 Week: 25	2,561 (99% content peer-reviewed journals)	Psychological, social, behavioural and health sciences.
WoS	1898 – 2015 Week: 27	More than 12,000 (high impact journals)	Medical sciences, social sciences, arts and humanities.

Journal selection and search

The International Journal of Therapeutic Communities is the only peer-reviewed journal in existence publishing articles related to this subject area. Previous reviews have conducted a hand search of the journal due to its comprehensive coverage on therapeutic communities and limited indexed availability online (Lees, *et al.,* 2004). The Planned Environments Therapy Trust (PETT) archives were visited and a hand search of the journal was undertaken to identify relevant studies from 1999 to the current date.

Data Extraction

For each study the following information was recorded: author, date of publication, demographics of the study population/s, percentage of follow up, definition and type of TC, interpersonal and offending risk outcome measures used, and key findings (Appendix B - Table 2). Meta-analysis was considered inappropriate due to the heterogeneous nature of findings collected across relevant studies with respect to outcome measures used.

Assessment of methodological quality

A combination of generic rating tools (e.g., Critical Appraisal Skills Programme

[CASP], University of Oxford, 2005; Newcastle-Ottawa Scale, Wells *et al*., 2010) in addition to methodological issues highlighted by previous reviews specific to therapeutic community research (Magor-Blatch *et al.,* 2014; Lees *et al.,* 2004) were integrated to develop a quality assessment tool for this review.

The quality assessment tool contained ten[1] questions that examined four potential sources of bias, including: participants, design bias, and assessment of outcomes. A rating out of three was awarded for each question and an overall score was achieved through summing the points awarded. A number of issues have been noted in regard to use of scoring systems within quality tools, particularly in their ability to provide a representatively weighted total quality value (Sanderson, Tatt & Higgly, 2007). This tool was used in combination with qualitative description to provide an accessible indication of study quality from which to weight research results.

Results

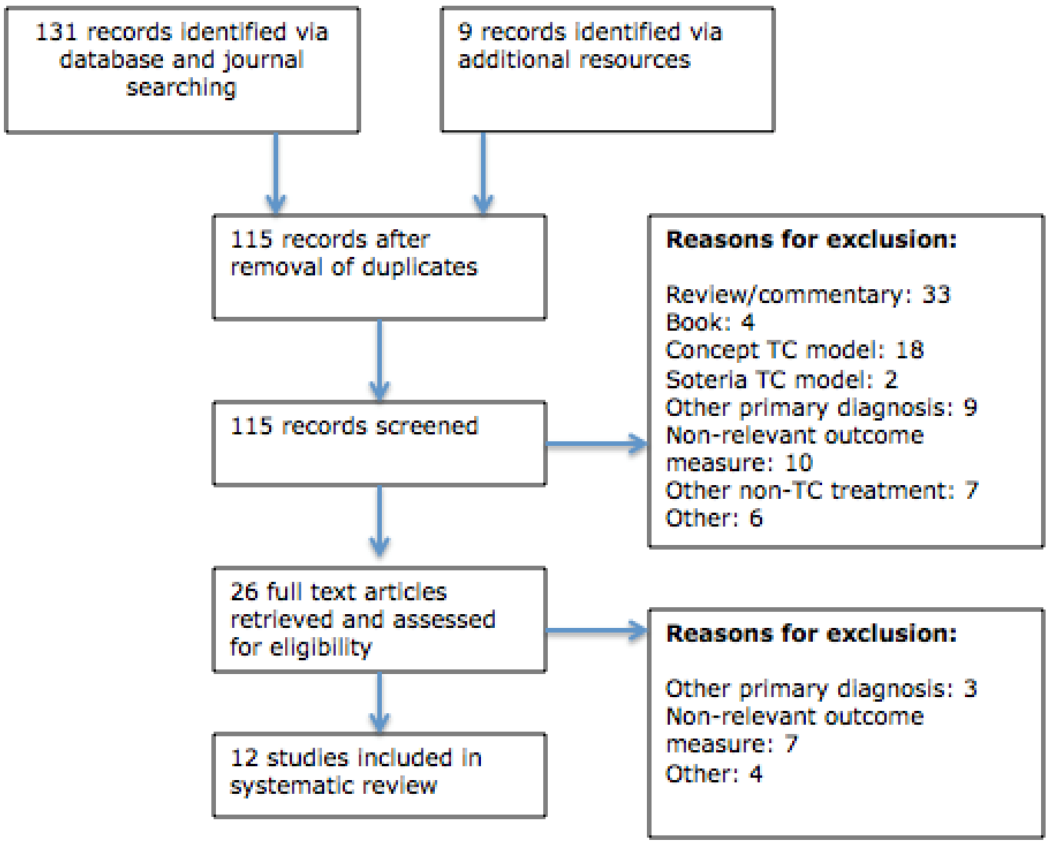
Selection

The database search retrieved 131 studies. While the journal search retrieved four additional studies, only three could be sourced. Six additional studies were found through reference trawling key articles giving a total sample of 115 studies after duplicate removal. All studies retrieved were screened to produce a sample of twelve studies for review. Three studies were pooled together as they measured the same populations at different time points (Chiesa *et al.,* 2002; Chiesa, Fonagy & Holmes, 2003; Chiesa *et al*., 2004).

This resulted in ten studies for review. Further detail regarding the process of study selection has been presented in a flow diagram as recommended by the PRISMA group (see Figure 1 below; Moher *et al.,* 2010).

Figure 1. PRISMA Diagram

[1] 1. Were exposed participants representative of the wider population of interest? 2. What was the quality of the control/comparison group? 3. Did the study control for differences in demographic characteristics between groups? 4. Did the study control for differences in important clinical variables between groups? 5. Did the study control for attrition? 6. Were assessors blinded? 7. Did the study report of reliability/validity of measures? 8. What was the quality of the measures used? 9. What was the length of follow up? 10. What proportion of participants were followed up?



A summary of the design and key results of the ten reviewed studies is given in Table 2 below.

Table 2 – Data extraction table

Primary author/year	Participants	% FU	Gender and mean age (years)	Treatment (max duration in months)	Outcome Measures (duration in months)	Key findings*
Barr (2010)	29 Attrition: 9	-	M/F 35	Four Day TCs (One day p/w) NR (median duration 51.5 weeks)	PDQ-4 CORE SFQ ZAN-BPD TAG A and D (or 12 months after A)	Small effects identified on PDQ (*d*= 0.27, CI -0.48 to 1.04) and researcher measure (ZAN-BPD; *d*=0.38, CI -1.46 to 2.23). Medium effects noted on CORE (*d*=0.62, CI -6.91 to 8.19) and SFQ (*d*=0.72, CI -0.40 to 1.87). Large effect size on clinical team report measure (TAG) outcome (*d*=1.48, CI 0.57 to 2.47).
Birtchnell (2009)	410 Attrition: 280 *Comparison group* MSU population 81 Attrition:	- 37	M 34 M 29	PTC (Accred) 18 MSU (non-TC) Up to 12	PROQ3 (A, 9, 18) PROQ2 (PT, 3, 9, and 12 [FU] – NR)	A – 18: PTC demonstrated medium effect (*d*=0.68, CI -1.54 to 2.90). A – 9: MSU demonstrated medium effect (*d=0*.65, CI -7.08 to 8.41). Large treatment effect size identified (*d*=2.42, CI 2.13 to 6.99) when PTC treatment compared with MSU at 9 months.

Chiesa (2004) **Chiesa (2003)** **Chiesa (2002)**	79 Attrition: 30 Step-down programme 58 Attrition: 13 TAU 73 Attrition: 24	96 100 86	M/F 32 M/F 32 M/F 35	Residential TC 10-11 Step-down programme 7 (IP TC) + 14 OP TAU 24	SCL90-R SAS GAS (A, 6, 12, 24)	Residential TC vs. TAU: Limited effects on GSI at 12, and 24 months (*d*=0.13, CI -0.03 to 0.28), small to medium effect on GAS at 12 (*d*=0.33, CI -1.19 to 1.87), and 24 months (*d*=0.56, CI -0.96 to 2.10) and medium to large effects on SAS at 12 (*d*=0.67, CI 0.62 to 0.73), and 24 months (*d*=1.00, CI 0.93 to 1.09). Step-down programe vs. residential TC: Small to medium treatment effect on GSI at 12 (*d*=0.33, CI 0.16 to 0.52) and 24 months (*d*=0.56 CI 0.16 to 0.52). GAS scores indicated small effects at 12 (*d*=0.48, CI -2.37 to 3.35) and 24 months (*d*=.41, CI -2.55 to 3.37) compared to residential TC. Medium treatment effects for step-down group compared to residential TC outcomes on SAS at 12 (*d*=0.50, CI 0.39 to 0.63) and 24 months (*d*=0.60, CI -0.51 to 0.71).
Jones (2013)	25 Attrition: 18 *Comparison group* SUN Project 37 Attrition: 22	- NR -	M/F 39 Up to 18 Drop in TC (Option to attend up to 4 times p/w) 6	Day TC (2-3 days p/w) CORE GAF HoNOS BSI SAS-SR IIP (A, 10, 18) ES 6		A-18: Large effect sizes noted in HoNOS (*d*=1.35, CI -1.51 to 4.42), GAF (*d*= 1.54, CI -4.33 to 7.65), and BSI (*d*= 2.68, CI 0.10 to 5.68). Small negative effects for SAS-SR (*d*=-0.21, -0.57 to 0.11), CORE (*d*= -0.27, CI -0.94 to 0.36), and IIP (*d* =-0.18, CI -0.58 to 0.20). ES: Large effect identified from participants perceived change in self-esteem post 6 months SUN membership across all subscales (*d*=3.17, CI 2.10 to 4.83).
McFetridge (2010)	116 Attrition: 76	34	F 32	Residential TC (Accred)	CORE (A, 5-80)	Large effect identified at FU on CORE scores for completers of therapy (*d*=1.93). Large effect (*d*=1.02), between completers and non-

				Up to 12		completers, with completers of therapy having lower total CORE scores. Gains evident at long term FU for completers of therapy at average of 5 years.
Morrissey (2014)	13 Attrition: 4	-	M 32	High secure LDTC 24	YSQ – SF IPDE Screen PCL-SV (A, 24) Seclusion rate (6 PT, 6, 12, 18, 24)	Large effect sizes on YSQ-SF subscales (*d*=0.89 to 1.17): Vulnerability to Harm (*d*=1.17, CI -0.18 to 2.66), Entitlement (*d*=1.13, CI 0.03 to 2.36), Emotional Inhibition (*d*=0.89, CI -0.66 to 2.56), Defectiveness/Shame (*d*=0.99, CI -0.73 to 2.81). Medium/small effects found for: Emotional deprivation (*d*=0.62, CI -1.30 to 2.62), Unrelenting Standards (*d*=0.57, CI -0.71 to 1.91), enmeshment (*d*=0.29, CI -1.06 to 1.67), subjugation (*d*=0.31, CI -0.65 to 1.31). IPDE: Large effect sizes found on paranoid (*d*=1.41, CI 0.82 to 2.2), schizoid (*d*=1.29, CI 0.73 to 2.01), and antisocial (*d*=1.11, CI 0.59 to 1.77) subscales. Medium and small effects for schizotypal (*d*=0.77, CI 0.03 to 1.67) and borderline (*d*=0.21, CI -0.79 to 1.24) scales. PCL-SV – Limited treatment effect (*d*=0.09, CI -1.32 to 1.51). Mean seclusion hours reduced by 90%; from 33 hours (0-6 months) to 2 hours (18-24 months).
Pearce (2008)	7 Attrition: 11	-	M/F NR	Mini TC – Community (5 hours per week) + step-down group 2	CORE (A, D; specific test periods NR)	CORE – Improved scores on all areas; functioning (79%), wellbeing (74%), problems (65%), risk (97%).

				months pre-discharge Up to 24		
Rawlinson (2010)	22 Attrition: 6	-	M/F 35	Day TC (4.5 days a week) Up to 18	CORE (A, D; specific test periods NR)	T-tests identified significant reductions in pre/post scores for unplanned/planned discharges (p<0.05) in all CORE subscales – improvements in wellbeing, functioning and reduction in risk and problems. Decrease of 0.5 or more between pre and post treatment scores indicated change was reliable and sustainable over time.
Shuker (2008)	291 Attrition: 119	-	M NR	PTC (Accred) Up to 36	EPS HDHQ (A, D; specific test periods NR) PBA NR	Improved scores on risk measures; EPS scales of impulsiveness, and psychoticism, and HDHQ scales (all at p<.001; effect sizes (d=0.5 to 0.9). Improvements on offending risk domains correlated with reduced risk within PBA assessments (effect sizes *(d=0.*2 to 0.6). RCI indicated only men who left after more than one year demonstrated clinically significant change on all risk measures.
Wilson (2014)	47 Attrition: NR	-	M 33	Medium secure TC (Accred) Up to 36	HCR – 20 VRS SCL90 – R (A, 12, 24, 36)	A to 36 months - HCR20: Limited effect (d=0.08, CI -0.64 to 0.81). VRS: Large effect on risk of violence indicated by dynamic item scores (d=0.88, CI -0.36 to 2.14). SCL90-R: GSI scores indicated a medium effect (d=0.75, CI 0.07 to 0.87).

Notes: * Results converted to Cohen's *d* effect sizes where possible and reported with 95% confidence intervals to enhance comparability of findings; % follow up in months, percentage of N with post-treatment outcome measures; A, outcome measured at admission; Accred, Accredited TC status by the Community of Communities; BSI, Brief Symptom Inventory (Derogatis & Melisaratos, 1983); CORE, Clinical Outcomes in Routine Evaluation questionnaire (Evans *et al.,* 2002); D, outcome measured at discharge; EPS, Eysenck Personality Scales (Eysenck & Eysenck, 1975); ES, Empowerment Scale (Rogers *et al.,* 1997); F, female; FU, follow up; GAF, Global Assessment of functioning (Hall, 1995); GAS, Global Assessment Scale (Endicott *et al.,*1976); GSI, Global Severity Index of SCL90-R (Derogatis, 1994); HCR-20, Historical Clinical Risk Management – 20 (Douglas *et al*., 2014); HDHQ, The Hostility and Direction of Hostility Questionnaire (Philip, 1969); HoNOS, Health of the National Outcome Scale (Wing, Beever & Curtis, 1998); Individual contrasts, comparison of baseline data with separate treatment time points; IPDE Screen, International Personality Disorder Examination (Loranger, 1997); IIP, Inventory of Interpersonal Problems (Horowitz *et al.,* 1988); IP, Inpatient treatment; LDTC, Learning Disability Therapeutic Community; M, male; MSU, Medium secure unit; PDQ-4, Personality Disorder Questionnaire – Four (Hyler, 1994); PROQ, Person's Relating to Others Questionnaire (Birtchnell, Falkowski & Steffert, 1992); NR, not reported; N.S, non-significant results; OP, Outpatient treatment; PBA, Parole board assessment; Post, post-treatment data; Pre, baseline data; PT, outcome measured pre-treatment; PTC, prison TC; RCI, Reliable Change Index; rMANOVA, repeated measures multivariate analysis of variance; SAS-SR, Social Adjustment Scale – Self Report (Weissman, 1999); SCL90 – R, The Symptom Checklist 90 – Revised (Derogatis, 1994); SFQ, Social Functioning Questionnaire (Tyrer *et al*., 2005); SH, self-harm; TAU, Treatment as usual – remains under community health team and referring clinician; Threshold Assessment Grid, TAG (Slade *et al.,* 2000); VRS, Violence Risk Scale (Wong & Gordon, 1999); YSQ-SF, Young Schema Questionnaire – Short Form (Young, 1998); ZAN-BPD, Zanarini Rating Scale for Borderline Personality Disorder (Zanarini, 2003).

Quality assessment results

Nine of the studies subjected to quality assessment achieved between 12-16 points out of a possible score of 30 (Barr *et al.,* 2010; Birtchnell *et al.,* 2009; Jones, Juett & Hill*,* 2013; McFetridge & Coakes, 2010; Morrissey & Taylor, 2014; Pearce & Haigh, 2008; Rawlinson & Bennett, 2010; Shuker & Newton, 2008; Wilson *et al.,* 2014). The low range of these scores suggests a high risk of bias within the majority of included studies for review.

One study scored 25, indicating a more robust study design and decreased risk of bias (Chiesa *et al.,* 2004). Study results are considered in relation to potential sources of bias examined by the quality assessment tool.

Participants

Representativeness of participant samples

The majority of participants were recruited via psychotherapy referral. However, it was unclear whether recruited samples in community TCs were then subject to TC member approval. Some studies harboured increased risk of sample bias. Morrissey & Taylor (2014) selected participants into the LDTC on a pragmatic basis and McFetridge & Coakes (2010) invited ex-clients to participate on a voluntary basis.

Further risk of bias existed within some studies due to small sample size (Barr *et al.,* 2010; Jones *et al.,* 2013; Morrissey & Taylor, 2014; Pearce & Haigh, 2008; Rawlinson & Bennett, 2010) and subjective (screening) assessments of PD (IPDE, PDQ – Barr *et al.,* 2010; Morrissey & Taylor, 2014; Rawlinson & Bennett, 2010; Shuker & Newton, 2008). In four studies, retested participants were more likely to have been in treatment longer (Barr *et al*., 2010; McFetridge & Coakes, 2010; Shuker & Newton, 2008; Wilson *et al.,* 2014).

Control/comparison group

Absence of a control group (e.g. waiting list control group) causes difficulty in determination of whether observed effects are specific to the target population (Meltzoff & Kornreich, 2008). No studies employed a control group. Seven of the studies reviewed had no comparison group (Barr *et al.,* 2010; McFetridge & Coakes, 2010; Morrissey & Taylor, 2014; Pearce & Haigh, 2008; Rawlinson & Bennett, 2010; Shuker & Newton, 2008; Wilson *et al.,* 2014).

In Chiesa et al.'s (2004) study all groups were subject to the same inclusion/exclusion criteria. Results from this study have greater specificity due to

increased comparability of participant groups, in addition to multiple group contrasts; two additional treatment groups (TAU, residential TC plus step down treatment).

Design bias

Controls for demographic characteristics

Demographic variables known to have a relationship with PD and treatment and offending outcomes include: age, gender, marital status, childhood abuse and neglect, socio-economic status indicators (employment and education), and forensic history (e.g. number of previous offences) (Chiesa *et al*., 2004; Miller & Brown, 2010; Wilson *et al.,* 2014). Aside from Chiesa et al. (2004), no other studies reported information for all of these variables. Studies that have not used a control or comparison group would be expected to report on demographic and clinical variables to enable comparison with other clinical studies. This remains the only way to establish whether the population under review is representative of both the clinical population and the general population. Of the three studies that included more than one group, one controlled for all relevant demographic variables reported (Chiesa *et al.,* 2004).

Controls for clinical variables

Clinical variables known to have a relationship with PD and treatment outcome include: PD comorbidity, DSM axis I/axis II comorbidity (e.g. depression and PD), degree of impulsivity, and psychopathy (Bateman & Fonagy, 2000; Harris *et al.,* 1994; Schilling *et al.,* 2012). Of the three studies with a comparison group, none reported information or controlled for all of these variables. Chiesa *et al.* (2004) reported PD comorbidity to be well matched across groups subsequent to statistical comparison, although a significant difference was noted in major depression status, which was controlled for in statistical analysis.

Controls for attrition

Attrition has been known to influence research results, increasing risk of sample bias, alongside reducing statistical power (Weiner, Schinka, & Velicer, 2003). Five studies incurred a drop out rate of ≥50% (59-72%) (Birtchnell et al., 2009; Jones *et al.,* 2013; McFetridge & Coakes, 2010; Pearce & Haigh, 2008) or attrition rates were not reported (Wilson *et al.,* 2014). Two studies retained data from lost participants pre-discharge via intention to treat (ITT) analysis (Chiesa *et al.*, 2004; Wilson *et al.*, 2014). Two studies excluded data from lost participants from pre-post comparisons of treatment effect (Barr *et al.,* 2010; Jones *et al.,* 2013).

Shuker & Newton (2008) included participants who did not complete treatment but were re-tested prior to leaving the prison TC.

Assessment

Six different measures were used to explore offending risk outcomes, while 15 distinct assessments were used to explore interpersonal outcomes, complicating comparison of results. Assessments of outcomes are considered via research setting (day/mini, residential and secure).

Blinding of assessors

Blinding outcome assessors reduces experimenter bias (Hróbjartsson *et al.,* 2012). Blinding was used in one residential TC study (Chiesa *et al.,* 2004), where independent researchers completed selected interpersonal measures (SAS-R and GAS). A proportion of tests were independently analysed by a psychiatrist blind to group allocation.

Standard of measures

Mini/day/residential TCs: Four out of six studies did not report on reliability or validity of measures used (Barr *et al.,* 2010; McFetridge & Coakes, 2010; Pearce & Haigh, 2008; Rawlinson & Bennett, 2010). No studies reported on the quality of measures in respect to evidenced reliability and validity with a PD population. Jones *et al.* (2013) reported all measures had established reliability and validity in assessment of treatment outcome, although provided no statistical information. Chiesa *et al.* (2004) reported solely on the inter-rater reliability of SAS-R and GAS measures (Interclass Coefficient 0.78, 0.79 respectively).

Forensic TCs: One study solely reported the test-retest reliability of an interpersonal measure used to be satisfactory (SCL90-R) (Wilson *et al.,* 2014). Birtchnell *et al.,* (2009) reported the PROQ to have established inter-rater reliability and construct validity, although these properties were demonstrated in a general psychotherapy sample. Shuker and Newton (2008) noted all measures utilized to have good predictive validity as risk factors for re-offending. Morrissey and Taylor (2014) reported the YSQ-SF and IPDE measures had not been validated for use with an Intellectual Disability population. However, the PCL-SV (offending risk measure) had established inter-rater reliability and predictive validity with this population.

Follow up

No studies based within mini, day or forensic TCs employed follow up measures for the treatment group (Barr *et al.,* 2010; Birtchnell *et al.,* 2009; Jones *et al.,* 2013; Pearce & Haigh, 2008; Rawlinson & Bennett, 2010; Shuker & Newton, 2008; Wilson *et al.,* 2014). This limited conclusions regarding the sustained efficacy of the interventions.

Residential TCs: The two studies based in residential TCs utilized follow up measures, although time of administration was asymmetrical, increasing risk of bias (Chiesa *et al.,* 2004; McFetridge & Coakes, 2010).

Interpersonal and offending risk outcomes

Participant demographics

Four studies based in forensic TCs included male participants only. Residential, day and mini TCs included a study of one female only sample, and five studies contained male and female participants. Females ranged from representing 35-88% within mixed samples. The mean age of participants in each study ranged from 29-39 years. Between 82-100% of participants had diagnoses or met screening criteria for PD.

Characteristics of TC programmes in included studies

All studies identified with traditional DTC philosophies, principles and practices, such as community meetings and flattened hierarchies. Eight studies provided clear programme details via reference to TC principles or practices, reporting at least three treatment components. Additional treatment elements were outlined by seven studies and included CBT, DBT, Cognitive Analytic Therapy (CAT), Mindfulness Based Therapy (MBT), art psychotherapy, transactional analysis, and psychodrama. Four studies (three forensic, one residential) were conducted in settings accredited by the Community of Communities, which provided a further level of fidelity regarding consistent implementation of TC principles.

Findings

It is recognized that some of the measures discussed below are measures of psychiatric symptomatology and are not primarily measures of interpersonal functioning. However, the majority of the clients in TCs have a diagnosis of BPD. Consequently, support from a TC intervention is often primarily sought due to difficulties with emotional regulation and relational functioning.

In addition, the main focus of TCs reside upon improving relational functioning. Therefore, any measure of distress is likely to reflect distress related to socio-emotional functioning and any sustained changes in distress could be inferred as related to changes in socio-emotional functioning.

In seven studies, selected tests measured a number of outcome areas and contained a limited number of items relevant to interpersonal outcomes (CORE-OM, GAF, GAS, HoNOS, SCL90-R, TAG). Use of these scales prevented meaningful interpretation of results in studies where interpersonal subscales were not reported

on (Jones *et al.,* 2013; McFetridge & Coakes, 2010; Wilson *et al.,* 2014). Consequently, findings are discussed in relation to studies that employed specific interpersonal measures and subscales and/or offending risk outcomes. Results are considered in accordance with methodological design of included studies: within subjects design (pre and post test analysis), and TC compared with another treatment group design.

Within subjects TC comparison

Interpersonal outcomes: While five studies focused on pre and post comparisons of interpersonal outcomes (Barr *et al.,* 2010; McFetridge & Coakes, 2010; Morrissey & Taylor, 2014; Pearce & Haigh 2003; Rawlinson & Bennett, 2010), only two studies (Barr *et al.,* 2010; Morrissey & Taylor, 2014) reported on interpersonal measures and subscales for which effect sizes could be calculated.

Morrissey and Taylor (2014) identified a number of large treatment effects after 24 months of LDTC treatment across YSQ-SF subscales: emotional inhibition, entitlement, vulnerability, defectiveness/shame (*d*=0.89 to 1.17), with medium and small effects on unrelenting standards (*d*=0.57), enmeshment (*d*=0.28), and subjugation subscales (*d*=0.31). A number of moderate to large effects were also noted on IPDE subscales: paranoid (*d*=1.41), schizoid (*d*=1.11), antisocial (*d*=1.11), schizotypal (*d*=0.77). Barr *et al.* (2010) identified small treatment effects indicating reduced levels of personality disorder symptomology (PDQ, d=0.27, CI -0.48 to 1.04; ZAN-BPD, d=0.38, CI -1.46 to 2.23) and a large effect for improved social functioning (*d*=0.72, CI -0.40 to 1.87) post 12 months of one-day DTC treatment. However, variability in confidence intervals on all three measures suggested level of improvement varied considerably within the sample.

Offending risk outcomes: Three forensic TC studies completed pre and post comparisons on offending risk outcomes (violence and reconviction risk), and provided mixed findings. Wilson *et al.* (2014) found a large treatment effect for risk of violence (VRS, *d*=0.88, CI -0.36 to 2.14) 36 months post treatment, while no effect was identified for violence risk based on HCR-20 scores (HCR20, *d*=0.08, CI -0.64 to 0.81).

Overlap in confidence intervals highlighted considerable variability in scores, and suggests VRS scores may not represent a valid or reliable effect. However, as attrition rate was not reported, calculations of effect sizes were made with the original number of participants, which may have biased results. In contrast, Morrissey and Taylor (2014) identified a 90% reduction in seclusion use after two years of treatment.

Shuker and Newton (2008) identified medium effects for PTC treatment on reconviction risk associated personality traits (EPS subscales; Impulsiveness, *d*=0.6, CI 0.4 to 0.8, Psychoticism, *d*=0.5, CI 0.2 to 0.7; Extrapunitive hostility subscales

[HDHQ], *d*= 0.5, CI 0.3 to 0.8). Small to medium effect sizes were demonstrated on reconviction risk measures, for men who were granted parole compared to those who were not; Impulsiveness (*d*=0.6, CI 0.4 to 0.9), Psychoticism, (*d*=0.4, CI 0.2 to 0.7), Extrapunitive Hostility (*d*=0.2, CI 0 to 0.5). A reliable change index (RCI), (Jacobson, Follette & Revenstorf, 1984), demonstrated clinically significant change on all risk measures was only achieved after treatment exposure of a year or more. Alternatively, Morrissey and Taylor (2014) identified a limited effect after 24 months of LDTC treatment on level of psychopathic traits (PCL-SV; *d*=0.09).

TC compared with another treatment group

Interpersonal outcomes: Three studies compared TC treatment with either one (Birtchnell *et al.,* 2009; Jones *et al.,* 2013), or two (Chiesa *et al.,* 2004) separate treatment groups. All studies focused on interpersonal outcomes.

Comparisons between treatment groups were not possible in one study due to small sample sizes (Jones *et al.,* 2013). The remaining two studies found partial evidence for residential (Chiesa *et al.,* 2004) and prison-based forensic (Birtchnell *et al.,* 2009) DTC treatment. Birtchnell *et al.,* (2009) found a large positive effect for PTC treatment when compared to an MSU sample on styles of relating with others at 9 months post treatment (PROQ; *d*= 2.42, CI 2.13 to 6.99). However, results may have been biased in regard to differences between samples at baseline. MSU participants demonstrated increased interpersonal difficulties in comparison to the PTC sample as indicated by mean pretreatment PROQ scores (MSU; 124.5, PTC; 51.7), which might explain wide variability highlighted in confidence intervals. Chiesa *et al.,* (2004) compared residential TC treatment with combination treatment (residential TC and step-down group) and a TAU group. Moderate to large effects were achieved by residential TC treatment on social adjustment (SAS-R) at 12 (*d*= 0.67) and 24 months (*d*=1.00) when compared to the TAU group.

Comparison of residential TC and combination post-treatment results demonstrated a medium effect for combined treatment on social adjustment at 12 months (*d*=0.50), which increased at 24 months (*d*=0.60), with participants exposed to shorter inpatient stay and an additional step down group achieving greater improvements.

Discussion

Overview

This review examined interpersonal and offending risk outcomes for individuals with PD following DTC treatment, with respect to how outcomes were measured, and whether treatment setting influenced outcomes. While evidence was provided for

improved interpersonal outcomes post DTC treatment in forensic and residential settings, evidence remained mixed for the efficacy of DTC treatment in day and mini TC settings. Evidence provided by one study with increased methodological rigour found combined treatment (shorter residential TC inpatient stay and step down treatment) was superior to residential treatment alone on interpersonal outcomes. Mixed results were also demonstrated for offending risk outcomes (risk of violence and reconviction) within forensic TCs. Nine out of ten studies available were of a low study quality generally and contained extensive methodological limitations varying from representativeness of participants to measures used and time periods of assessment, all of which were likely to have biased results. The conclusions of this review are therefore limited due to the limitations of research available.

Clinical and research implications

The mixed evidence for DTCs demonstrated in this review lie in contrast to that of a previous meta-analytic review on the efficacy of DTC treatment for PD in secure and non-secure settings (Lees *et al.,* 2004). A large proportion of studies included in this review harboured numerous methodological limitations, (e.g. limited use of blinding, randomization, follow up periods, control or comparison groups) and samples were often drawn from highly varied populations (e.g. from prison to high secure in forensic TCs).

It is therefore possible mixed findings on interpersonal and offending risk outcomes are due to the methodological and design limitations of available studies and varied nature of samples between and within studies, as opposed to the limitations of DTC treatment for PD. Limited evidence of reduced offending risk post DTC treatment may be due to the high number of static items located on risk assessments (Gendreau *et al.,* 1996).

The majority of DTC studies assessing offending risk post-treatment utilized risk assessments with an increased number of static (past or historical risk) versus dynamic items (PCL-SV, HCR-20), limiting the validity of many of the studies and the literature as a whole.

In consideration of findings from a residential TC study with heightened methodological rigour, results lend partial support to TC theory, in respect to its aim to support individuals to develop adaptive styles of interpersonally relating (Haigh, 2013). However, post discharge from inpatient settings, it would seem interpersonal outcomes are enhanced by additional treatment in the community, for example, in the form of a step-down group. This finding suggests some difficulty in the translatability of interpersonal skills from inpatient TC settings to wider society and relates to a previous argument on the ongoing conflict between rehabilitation (preparation for the outside world) and psychotherapy within TCs (Campling, 2001).

It is possible step-down groups are better placed to achieve a balance between these two goals via continued facilitation of psychotherapy treatment while individuals are grounded in life outside of hospital, enabling application of psychotherapeutic skills to everyday life. Lees et al. (2004) did not compare outcomes of DTC treatment with a combination intervention such as a step-down group. Consequently, the sustained efficacy of DTC treatment demonstrated by previous studies for PD populations may be smaller than previously concluded.

Surprisingly, DTC research appears to have made limited progress in consideration of previous research recommendations outlined by Lees *et al.* (2004). Attrition rates continued to remain problematic, with a limited number of studies including measures to control for this, such as ITT. While a limited number of studies included treatment comparison groups, no control groups were utilised, despite recommendations for incorporation of waiting list controls (Warren *et al.,* 2003). A number of studies excluded follow up measures. Where used, follow up periods were conducted at different times between groups and of limited duration, preventing objective exploration of the sustainability of treatment outcomes. DTC treatment was defined in sufficient detail across studies, although description of selection procedures were vague in some cases, with no indication of whether participants were selected by community members post psychotherapy referral.

This review controlled for heterogeneity of client samples by stipulating inclusion criteria of participants with PDs via diagnosis or screening instruments. However, this criterion limited the number of studies applicable for review considerably, indicating heterogeneity of client samples to remain an ongoing issue in current research. As DTCs are most often considered for individuals with diagnostic complexity, decreased homogeneity within study samples may reflect the limitations of a diagnostic approach for complex mental health difficulties, including PD, as opposed to poor quality research designs (Maj, 2005).

Difficulties in neat diagnostic conceptualization of DTC client groups may also account for the varied measures used within current research in the assessment of treatment outcomes, due to heterogonous and complex nature of client groups admitted.

In light of the diverse populations treated by TCs and the complexity of the treatment process, it would seem that a positivist approach (focusing on 'what works' via objective measurement of observable phenomena), usually applied to manualised therapies, may fall short of capturing the essence of fundamental TC features (Haigh, 2014). Instead, a more flexible and pragmatic research design may required to identify 'what matters' (Haigh, 2005, p. 5), such as qualitative or mixed-methods approach. Future research reviews could be completed on qualitative/mixed methods studies via a mixed methods research review - a combination qualitative meta-synthesis and meta-analysis/narrative synthesis of quantitative data, as appropriate.

Future research should consider adopting a focus on processes inherent within TCs to develop an evidence-based understanding of its most effective components so we might enhance these aspects (Aslan & Yates, 2015; Magor *et al.,* 2014; Veale *et al.,* 2014). As highlighted by Aslan & Yates (2015), in light of the continuous modification of TCs and ongoing therapeutic integration of different theoretical approaches, gaining a deeper understanding of how TCs work would also enable future clinicians to approach TC modification with an understanding that would guard against undermining the core integrity of the model.

Limitations

A strength of the study is that it provided an evaluation of the current quality of DTC outcome research across a number of settings (day/mini, residential, forensic), and recommendations for the direction of future research. This review should also be interpreted with reference to its limitations. As participants were included from a range of settings, this compromised the depth of comparisons able to be made between studies. However, this was deemed necessary due to the limited published literature currently available in the area. The study specifically explored DTC research in peer-reviewed journals, which may have rendered the review to risk of publication bias.

Future reviewers could examine findings of outcome studies located in grey literature and employ specific interpersonal measures in future outcome studies. However, it would seem a focus on processes within TCs would lead to a more refined understanding of key therapeutic processes within the treatment modality and better inform clinical practice.

While use of effect sizes enabled comparison of results between studies, non-normal sample distributions were calculated on the basis of the mean, as the median value of sample data was neither reported nor amenable to calculation from information provided. This method may have distorted certain effect size estimates.

Conclusions

Due to the low quality of studies conducted and their methodological limitations, there remains insufficient evidence to determine interpersonal and offending risk outcomes for individuals with PD following DTC treatment, particularly in regard to whether these factors are influenced by treatment setting. There is some evidence to suggest that while residential TCs are effective in supporting a community inpatient population at an interpersonal level, treatment effects are enhanced when hospital stay is reduced and combined with a post discharge step down group in the community. Further studies pursuing qualitative exploration of important TC processes from staff and client perspectives are paramount to increase understanding of DTC efficacy in the treatment of the notoriously heterogeneous diagnostic group that is PD.

References

Aslan, L. and Yates, R. (2015), "Exploring the black box of therapeutic community (TC) methodology and the subjective experiences of residents within TC structures", *Therapeutic Communities: The International Journal of Therapeutic Communities*, Vol. 36, No. 2.

Banerjee, P., Gibbon, S. and Huband, N. (2009), "Assessment of personality disorder", *Advances in Psychiatric Treatment*, Vol. 15, No. 5, pp. 389-397.

Barr, W., Kirkcaldy, A., Horne, A., Hodge, S., Hellin, K., and Gopfert, M. (2010), "Quantitative findings from a mixed methods evaluation of once-weekly therapeutic community day services for people with personality disorder", *Journal of Mental Health*, Vol. 19, No. 5, pp. 412-421.

Bateman, A. and Fonagy, P. (2000), "Effectiveness of psychotherapeutic treatment of personality disorder", *British Journal of Psychiatry*, Vol. 177, pp. 138-143.

Bateman, A. and Tyrer, P. (2004), "Psychological treatment for personality disorders", *Clinical Psychology and Psychotherapy*, Vol. 10, No. 5, pp. 378-388.

Birtchnell, J., Falkowski, J. and Steffert, B. (1992), "The negative relating of depressed patients: A new approach", *Journal of Affective Disorders*, Vol. 24, pp. 165-176.

Birtchnell, J., Shuker, R., Newberry, M. and Duggan, C. (2009), "An assessment of change in two male forensic therapy samples using the Person's Relating to Others Questionnaire (PROQ)", *The Journal of Forensic Psychiatry and Psychology*, Vol. 20 No.3, pp. 387-407.

Campling, P. (2001), "Therapeutic Communities", *Advances in Psychiatric Treatment*, Vol. 7, pp. 365-374.

Chiesa, M., Bateman, A., Wilberg. T. and Friis, S. (2002), "Patients' characteristics, outcome and cost-benefit of hospital-based treatment for patients with personality disorder: A comparison of three different programmes", *Psychology and Psychotherapy: Theory, Research and Practice*, Vol. 75, No. 4, pp. 381-392.

Chiesa, M., Fonagy, P. and Holmes, J. (2003), "When less is more: An exploration of psychoanalytically oriented hospital-based treatment for severe personality disorder", *International Journal of Psychoanalysis*, Vol. 84 No. 3, pp. 637-650.

Chiesa, M., Fonagy, P., Holmes, J. and Drahorad, C. (2004), "Residential versus community treatment of personality disorders: A comparative study of three treatment programmes", *American Journal of Psychiatry*, Vol. 17 No. 8, pp. 1463-1470.

Coid, J. *et al.* (2007). *Predicting and understanding risk of reoffending: The prisoner cohort study*, available at: http://www.crim.cam.ac.uk/people/academic_research/david_farrington/priscomoj.pdf (accessed 26 July 2015).

Derogatis, L. (1994). SCL-90-R: *Administration, scoring and procedures manual*. Minneapolis, MN: National Computer Systems.

Derogatis, L. and Melisaratos, N. (1983), "The Brief Symptom Inventory: An introductory report", *Psychological Medicine*, Vol. 13, pp. 595-605.

Douglas, K., Hart, S., Webster, C., Belfrage, H., Guy, L. and Wilson, C. (2014), "Historical-Clinical-Risk Management-20 Version 3 (HCR-20 v3): Development and overview", *International Journal of Forensic Mental Health*, Vol 13, No. 2, pp. 93-108.

Endicott, J., Spitzer, R., Fleiss, J. and Cohen, J. (1976), "The Global Assessment Scale: A procedure for measuring overall severity of psychiatric disturbance", *Archives of General Psychiatry*, Vol. 33, No. 6, pp. 766-771.

Evans, C., Connell, J., Barkham, M., Margison, F., McGrath, G., Mellor-Clark, J., and Audin, K. (2002), "Towards a standardized brief outcome measure: Psychometric properties and utility of the CORE-OM", *The British Journal of Psychiatry*, Vol. 180, No. 1, pp. 51-60.

Eysenck, H. and Eysenck, S. (1975). *Manual of the Eysenck Personality Questionnaire*. Kent: Hodder and Stoughton.

Fok, M., Stewart, R., Hayes, R. and Moran, P. (2014), "Predictors of natural and unnatural mortality among patients with personality disorder: Evidence from a large UK case register", *PLOS*, Vol. 9, No. 7, pp. 1-9.

Gendreau, P., Little, T. and Goggin, C. (1996), "A meta-analysis of the predictors of adult offender recidivism: what works!", *Criminology*, Vol. 34, pp. 557-607.

Haigh, R. (2014), "What is happening in the world of therapeutic communities?", *European Journal of Psychotherapy and Counselling,* Vol. 16, No. 1, pp. 74-76.

Haigh, R. (2013), "The quintessence of a therapeutic environment", *Therapeutic Communities,* Vol. 34, No. 1, pp. 6-15.

Haigh, R. (2005), "The trouble with modernization: We need better relationships, not policies and procedures", *The Mental Health Review*, Vol. 10, No. 3, pp. 3-7.

Haigh, R. (2002), Therapeutic community research: past, present and future, *Psychiatric Bulletin*, Vol. 26, pp. 65-68.

Haigh, R. and Tucker, S. (2004), "Democratic development of standards: the community of communities--a quality network of therapeutic communities", *The Psychiatric quarterly,* Vol. 75, No. 3, pp. 263.

Hall, R. (1995), "Global Assessment of Functioning: A modified scale", *Psychosomatics*, Vol. 36, No. 3, pp. 267-275.

Harris, G., Rice, M., and Cormier, C. (1994), "Psychopaths: Is a therapeutic community therapeutic?" *Therapeutic Communites*, Vol. 15 No. 4, pp. 283-299.

Heede *et al.* (2009), "Psychodynamic milieu therapy and changes in personality – what is the connection?" *Journal of Child Psychotherapy*, Vol. 35, No.3, pp.276-289.

Hopwood, C., Wright, A., Ansell, E. and Pincus, A. (2013), "The interpersonal core of personality pathology", *Journal of Personality Disorders*, Vol. 27, No.3, pp. 270-295.

Horowitz, L., Rosenberg, S., Baer, B., Ureno, G., & Villasenor, V. (1988), "Inventory of interpersonal problems: Psychometric properties and clinical applications", *Journal of Consulting and Clinical Psychology,* Vol. 56, pp. 885-892.

Hrobjartsson et al. (2012), "Observer bias in randomized controlled trials with binary outcomes: Systematic review of trials with both blinded and non-blinded outcome assessors", *The British Medical Journal,* Vol. 185, pp. 201-211.

Hyler, S. (1994). *Personality Diagnostic Questionnaire – 4 +.* New York: New York State Psychiatric Institute.

Jacobson, N. S., Follette, W. C. and Revenstorf, D. (1984), "Psychotherapy outcome research: methods for reporting variability and evaluating clinical significance", *Behavior Therapy,* Vol.15, pp. 336-352.

Jones, M. (1952), *Social Psychiatry: A study of therapeutic communities*, London: Tavistock Publications.

Jones, B., Juett, G. and Hill, N. (2013), "Initial outcomes of a therapeutic community- based outpatient programme in the management of personality disorder", *Therapeutic Communities,* Vol. 34, No. 1, pp. 41-52.

Kennard, D. (2004), "The therapeutic community as an adaptable treatment modality across different settings", *The Psychiatric quarterly,* Vol. 75, No. 3, pp. 295.

Kennard, D. (1998). *Introduction to Therapeutic Communities*. London: Jessica Kingsley.

Kennard, D. and Lees, J. (2012), "A checklist of standards for democratic therapeutic communities", *Therapeutic Communities,* Vol. 33, No. 2-3, pp. 117-123.

Kristman, V., Manno, M. and Cote, P. (2004), "Loss to follow up in cohort studies: How much is too much?" *European Journal of Epidemiology,* Vol. 19 No. 8, pp. 751-760.

Lees, J., Manning, N. and Rawlings, B. (2004), "A Culture of Enquiry: Research Evidence and the Therapeutic Community", *Psychiatric Quarterly,* Vol. 75, No. 3, pp. 279-294.

Lees, J., Manning, N. and Rawlings, B. (1999), "*Therapeutic Community Effectiveness: A systematic international review of therapeutic community treatment for people with personality disorders and mentally disordered offenders"*. University of York, York.

Loranger, A., Janca, A. and Sartorius, N. (1997). *Assessment and diagnosis of personality disorders: the ICD-10 international personality disorder examination (IPDE).* Cambridge: Cambridge University.

Magor-Blatch, L., Bhullar, N., Thomson, B. and Thorsteinsson, E. (2014), "A systematic review of studies examining effectiveness of therapeutic communities", *Therapeutic Communities,* Vol. 35, No. 4, pp. 168-184.

Madan, A. and Fowler, C. (2015), "Consistency and coherence in treatment outcome measures for borderline personality disorder", *Borderline Personality Disorder and Emotion Dysregulation*, Vol. 2 No. 1, pp. 1-4.

Maj, M. (2005), "Psychiatric comorbidity: An artefact of current diagnostic systems?" *The British Journal of Psychiatry*, Vol. 186, No. 3, pp. 182-184.

McFetridge, M. and Coakes, J. (2010), "The longer-term clinical outcomes of a DBT-informed residential therapeutic community; An evaluation and reunion", *Therapeutic Communities,* Vol. 31 No.4, pp. 406-416.

Meltzoff, J. and Kornreich, M. (2008), *Research in Psychotherapy*. New York: Transaction Publishers.

Miller, S. and Brown, J. (2010), "*HMP Dovegate's therapeutic community: an analysis of reconviction data*", *Therapeutic Communities*, Vol. 31 No. 1, pp. 62- 74.

Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group (2010), "Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement", *International Journal of Surgery*, Vol. 8 No. 5, pp. 336-341.

Morrissey, C. and Taylor, J. (2014), "Changes in Personality Disorder Traits Following 2 Years of Treatment in a Secure Therapeutic Community Milieu", *Journal of Mental Health Research in Intellectual Disabilities,* Vol. 7, No. 4, pp. 323-336.

National Institute for Health and Clinical Excellence (2009). *Antisocial personality disorder: Treatment, management and prevention*. CG77, Manchester: National Collaborating Centre for Mental Health.

National Institute for Health and Clinical Excellence (2010). *Borderline personality disorder: Treatment and management.* CG78, Manchester: National Collaborating Centre for Mental Health.

Newberry, M. (2010), "A synthesis of outcome research at Grendon Therapeutic Community prison", *Therapeutic Communities,* Vol. 31, No. 4, pp. 356-371.

Pearce, S. and Haigh, R. (2008), "Mini therapeutic communities – a new development in the United Kingdom", *Therapeutic Communities*, Vol. 29 No. 2, pp. 111-124.

Pearce, S. and Autrique, M. (2010), "On the need for randomised trials of therapeutic community approaches", *Therapeutic Communities*, Vol. 31, pp. 338–355.

Philip, A. (1969), "The development and use of the hostility and direction of hostility questionnaire", *Journal of Psychosomatic Research*, Vol. 13, pp. 283-287.

Polden, J. (2010), "Behind locked doors: An exploration of therapeutic processes within a prison therapeutic community", *British Journal of Psychotherapy,* Vol. 26, No. 4, pp. 502-521.

Rapoport, R. (1960). *Community as doctor: New perspectives on a therapeutic community.* London:Tavistock Publications, London.

Rawlings, B. (1999), "Therapeutic communities in prisons", *Policy And Politics; Policy Polit.,* Vol. 27, No. 1, pp. 97-111.

Rawlinson, D. and Bennett, C. (2010), "Evaluating the effectiveness of treatment in the therapeutic community: CORE outcomes and health care contacts", *Therapeutic Communities*, Vol. 31, No. 4, pp. 393-405.

Rogers, E., Chamberlin, J., Ellison, M. and Crean, T. (1997), "A consumer-constructed scale to measure empowerment among users of mental health services", *Psychiatric Services*, Vol. 48, No. 8, pp. 1042-1047.

Rutter, D. and Tyrer, P. (2003), "The value of therapeutic communities in the treatment of personality disorder: A suitable place for treatment?" *Journal of Psychiatric Practice*, Vol. 9, No.4, pp. 291-302.

Sanderson, S., Tatt, I. and Higgins, J. (2007), "Tools for assessing quality and susceptibility to bias in observational studies in epidemiology: a systematic review and annotated bibliography", *International Journal of Epidemiology*, Vol. 36, No. 3, pp. 666-676.

Schilling, M. *et al.* (2012), Normal mind-reading capacity but higher response confidence in borderline personality disorder patients, *Psychiatry and Clinical Neurosciences*, Vol. 66, pp. 322-327.

Shuker, R. (2010), "Forensic therapeutic communities: A critique of treatment model and evidence base", *The Howard Journal,* Vol. 49 No. 5, pp. 463-477.

Shuker, R. and Newberry, M. (2010). Changes in Interpersonal Relating Following Therapeutic Community Treatment at HMP Grendon. In: Shuker, R. and Sullivan, E. (eds.). *Grendon and the emergence of forensic therapeutic communities: Developments in research and practice.* Chichester, West Sussex: John Wiley & Sons.

Shuker, R. and Newton, M. (2008), "Treatment outcome following intervention in a prison- based therapeutic community: a study of the relationship between reduction in criminogenic risk and improved psychological well- being", *British Journal of Forensic Practice,* Vol. 10, No. 3, pp. 33-44.

Slade, M., Powell, R., Rosen, A. and Strathdee, G. (2000), "Threshold Assessment Grid (TAG): The development of a valid and brief scale to assess the severity of mental illness", *Social Psychiatry and Psychiatric Epidemiology*, Vol. 35, pp. 78-85.

Taylor, J., Crowther, S. and Bryant, C. (2015), "Therapeutic communities for people with intellectual disability and complex needs", *Adv Mental Hlth Intell Disabil,* Vol. 9, No. 3, pp. 124-131.

Tyrer, P., Nur, U., Crawford, M., Karlsen, S., McLean, C., Rao, B. and Johnson, T. (2005), "The Social Functioning Questionnaire: a rapid and robust measure of perceived functioning", *International Journal of Social Psychiatry*, Vol. 51, No., 3, pp. 265-275.

University of Oxford. (2005). *Critical Appraisal Skills Programme*. Available at: http://www.cebm.net/index.aspx?o=1157 (accessed 26 July 2015).

Vandevelde, S., Broekaert, E., Yates, R. and Kooyman, M. (2004), "*The development of therapeutic community establishments: A comparative retrospective account of the 'democratic' Maxwell Jones TC and the 'hierarchical' concept-based TC in prison",* (Unpublished Manuscript), University of Ghent, Belgium.

Veale, D., Gilbert, P., Wheatley, J. and Naismith, I. (2014), "A New Therapeutic Community: Development of a Compassion-Focussed and Contextual Behavioural Environment'", *Clinical Psychology & Psychotherapy*, Vol. 22, No. 4, pp. 285-303.

Warren, F. *et al*. (2003). *Review of Treatment for Severe Personality Disorder*. London: Home Office Online Report. Available at: http://www.floridatac.com/files/document/rdsolr3003.pdf (accessed 26 July 2015).

Wells, G. *et al.* (2010). *The Newcastle-Ottawa Scale (NOS) for assessing the quality of nonrandomised studies in meta-analyses*. Available at: http://www.ohri.ca/programs/clinical_epidemiology /oxford.asp. (accessed 26 July 2015).

Weiner, I., Schinka, J. and Velicer, W. (2003). *Handbook of psychology: Research methods in psychology*. 2ND Ed. New Jersey: John Wiley & Sons.

Weissman, M. (1999). *Social Adjustment Scale – Self-report (SAS-SR) User's Manual*. New York: Multi-Health Systems, Inc.

Widiger, T. (2012). *The Oxford Handbook of Personality Disorders*. Oxford University Press: New York

Wilson, K., Freestone, M., Taylor, C., Blazey, F. and Hardman, F. (2014), "Effectiveness of modified therapeutic community treatment within a medium-secure service for personality-disordered offenders", *The Journal of Forensic Psychiatry & Psychology,* Vol. 25, No. 3, pp. 243-261.

Wing, J., Beevor, A. and Curtis, R. (1998), "Health of the Nation Outcome Scales (Ho NOS)", *British Journal of Psychiatry*, Vol. 172, pp. 11-18.

Wong, S. and Gordon, A. (1999), *Manual for the Violence Risk Scale*. University of Saskatchewan: Canada.

Young, F. (1998). *The Young Schema Questionnaire: Short Form*. Available at: http://www.schematherapy.com/id55.htm (accessed 9 January 2016).

Zanarini, M. (2003), "Zanarini rating scale for borderline personality disorder (ZAN-BPD): A continuous measure of DSM-IV borderline psychopathology", *Journal of Personality Disorders*, Vol. 17, No. 3, pp. 233-242.

This is the final version submitted for publication in *Therapeutic Communities* prior to the publisher's edits (1st March 2016)

This is the final version submitted for publication in *Therapeutic Communities* prior to the publisher's edits (1st March 2016)