Evaluating and predicting therapy outcomes for patients at-risk of harming themselves

Katie Gregson-Curtis

Submitted for the Doctorate of Clinical Psychology

University of Sheffield

May 2017

Declaration

I confirm that this thesis has not been submitted for any other degree or to any other institution.

Word Count

Literature review	7,997 excluding references and appendices
	11,346 including references
	11,390 including references and appendices
Research report	11,998 excluding references and appendices
	14,814 including references
	18,269 including references and appendices
Total	19,995 excluding references and appendices
	29,659 including references and appendices
	29,906 including references, appendices, and overall abstract

Overall Abstract

Literature review: This review systematically evaluated evidence for the effectiveness of psychotherapies in improving psychological wellbeing among adults with recent self-injury. Search terms were used to identify studies published between January 1990 and February 2016 from the MEDLINE, CINAHL, and PsycINFO databases, and forward and backward reference searches were undertaken. Nineteen studies were reviewed. The most consistent treatment effects emerged for problem-solving therapy and emotion regulation group therapy on measures of mental health symptoms and hypothesised mechanisms of change. Treatment effects for other therapies were inconsistent or limited in breadth. It was concluded that psychotherapy can facilitate improvements in psychological wellbeing among adults with recent self-injury. Study limitations and recommendations for clinical practice and research were discussed.

Research report: The main aim of this study was to identify predictors of reliable and clinically significant improvement (RCSI) in therapy among patients "at-risk" of self-injury. A logistic multilevel modelling analysis was conducted using a national practice-based dataset. The sample included 4,976 patients who were treated by 81 therapists. Pre-post therapy scores on a measure of distress indicated that approximately half of the sample achieved RCSI. Lower pre-therapy distress, higher motivation, medication reductions, and being in employment or education predicted significantly greater odds of meeting criteria for RCSI. After taking these variables into account, therapist effects accounted for approximately 10-13% of the variance in outcomes. Some therapists were significantly more, or less, effective than average. Limitations of the study and recommendations for clinical practice and research were discussed.

Acknowledgements

Firstly, I would like to thank the patients, therapists, and service user groups who contributed information for the current study. I would also like to thank Professor Michael Barkham and Dave Saxon for their advice and guidance in executing this project. Further thanks go to the Sheffield DClinPsy course team and, specifically, to my tutors Dr Georgina Rowse and Sara Dennis, for their support throughout the duration of my training. Finally, a special thank you to my family, friends, and fellow trainees, for their help, patience, and encouragement.

List of Contents

Literature review

A systematic review of psychological wellbeing outcomes in psychotherapy with adults who self-injure

Abstract

Objectives: This systematic review evaluated evidence for the effectiveness of psychotherapies in improving psychological wellbeing outcomes among adults with recent self-injury.

Methods: Search terms related to self-injury and psychotherapy were used to identify articles published between January 1990 and February 2016 from the MEDLINE, CINAHL, and PsycINFO databases. Forward and backward reference searches identified additional studies.

Results: Nineteen studies with heterogeneous methods met inclusion criteria. Overall, study quality was fair. Consistent treatment effects emerged for problem-solving therapy and emotion regulation group therapy on mental health outcomes and hypothesised mechanisms of change. Treatment effects for other cognitive-behavioural therapies, psychodynamic-interpersonal therapy, and dialectical behaviour therapy were inconsistent or limited in breadth.

Conclusions: It was concluded that psychotherapy can facilitate improvements in psychological wellbeing among adults with recent self-injury.

Practitioner Points

Clinical Implications

- Psychotherapies, particularly problem-solving therapy and emotion regulation group therapy, may improve psychological wellbeing among adults with recent self-injury.
- Assessments of problem-solving and emotion regulation could inform decisions about appropriate therapies for patients.

Limitations

- The results may be less representative of men, and of people with schizophrenia, substance misuse difficulties, cognitive impairments, and developmental

differences.

- Most studies were randomised controlled trials with strict inclusion criteria. More evaluations of psychotherapy's effectiveness in routine practice are needed.

Globally, approximately one person dies each forty seconds by suicide (World Health Organisation [WHO], 2014). One of the strongest risk factors for suicide is a history of self-injury (Hawton et al., 2015; Zahl & Hawton, 2004). Self-injury is common. Between 4-6% of community-dwelling adults report previous self-injury (Briere & Gil, 1998; Klonsky, 2011) and these estimates increase to 12-80% in clinical populations (Jacobson, Muehlenkamp, Miller, & Turner, 2008; Washburn et al., 2012). Furthermore, 15-25% of people who attend hospital with self-injury re-present within one year (Carroll, Metcalfe, & Gunnell, 2014) and the repetition rate is higher among people with mental health difficulties (Larkin, di Blasi, & Arensman, 2014). Therefore, self-injury represents a significant public health concern.

Despite its prevalence, few established interventions for self-injury exist. Global and national policies suggest that one approach to reduce self-injury is psychotherapy (Department of Health, 2012; WHO, 2014). However, the effectiveness of different psychotherapies is unclear. The National Institute for Health and Care Excellence (NICE; 2011) recommends that people who self-injure be offered 3 to 12 sessions which may include elements of cognitive-behavioural therapy (CBT), psychodynamic therapy, or problem-solving therapy (PST). The lack of clear direction in this guidance reflects the conflicting evidence base. Whilst several reviews have suggested promising results for the effectiveness of CBT, psychodynamic therapy, and PST in reducing self-injury (Comtois & Linehan, 2006; Daigle, Pouliot, Chagnon, Greenfield, & Mishara, 2011), others have reported null treatment effects (Crawford, Thomas, Khan, & Kulinskaya, 2007; Hawton et al., 1999).

Comparisons of findings within and between reviews are complicated by the lack of conceptual clarity regarding self-injury. Few reviews have distinguished between suicidal self-injury (SSI) and non-suicidal self-injury (NSSI). SSI and NSSI differ in the presence, or lack, of suicidal intent. It has also been argued that SSI and

NSSI represent distinct difficulties that differ in prevalence, function, and frequency (Butler & Malone, 2013; Nock, 2009). Accordingly, a specific NSSI disorder was recently proposed as a potential diagnosis warranting further investigation (American Psychiatric Association, 2013). Research suggests that, among adults who self-injure, those meeting the NSSI disorder criteria report more frequent self-injury, higher impairment, and better responses to psychotherapy than those not meeting the NSSI disorder criteria (Zetterqvist, 2015). Therefore, SSI and NSSI may represent distinct needs which warrant distinct interventions.

Several systematic reviews have differentiated between SSI and NSSI. Turner, Austin, and Chapman (2014) reviewed forty trials of psychological and pharmacological interventions and concluded that dialectical behaviour therapy (DBT), emotion regulation group therapy (ERGT), manual-assisted cognitive therapy (MACT), and dynamic deconstructive therapy appeared promising for reducing NSSI. However, these conclusions were tentative. Only two studies investigated ERGT, and only one study each investigated MACT and dynamic deconstructive therapy. Many studies were excluded due to the review's focus on NSSI, which suggested that few existing studies had differentiated self-injury presentations. Ougrin, Tranah, Stahl, Moran, and Asarnow (2015) reviewed 19 randomised controlled trials (RCTs) of psychosocial interventions with adolescents. When studies of SSI and NSSI were combined, a significant treatment effect emerged on reductions in self-injury. DBT, CBT, and mentalization-based therapy had the largest effect sizes. Treatment effects were nonsignificant when studies of SSI and NSSI were evaluated separately. However, this result may have been impacted by the small number of studies producing low statistical power.

Therefore, research suggests that psychotherapy can facilitate reductions in self-injury. However, results are inconsistent and study comparisons are hampered by differing conceptualisations of self-injury. Furthermore, whilst reducing self-injury is

important, few reviews have investigated broader psychological wellbeing outcomes in psychotherapy for this population. NICE (2004, 2011) advocate that interventions should aim to improve patients' underlying difficulties, quality of life, and mental health, rather than solely reducing self-injury. People who self-injure are likely to encounter other stressful experiences. Indeed, in Madge et al.'s (2011) cross-cultural study of 30,477 adolescents, higher self-injury severity was significantly associated with poorer mental health and experiencing more negative life events.

Many theoretical models attempt to explain the relationship between experiencing stress and engaging in self-injury (see Barzilay-Levkowitz & Apter, 2014). Across these models, three risk factors consistently emerge: maladaptive cognitions, emotion regulation difficulties, and problem-solving deficits (Beck, Brown, Berchick, Stewart, & Steer, 1990; Gratz & Roemer, 2004; Linehan, 1993; Rudd, 2006; Williams & Pollock, 2001). These factors affect individuals' interpretations of events and ability to apply effective coping strategies. This can lead to feelings of hopelessness, thereby enhancing the likelihood of self-injury.

Direct empirical tests of models are rare. Indirect tests support that, compared to people who do not self-injure, people who self-injure report more maladaptive schemas related to social isolation and emotion inhibition (Lewis, Lumley, & Grunberg, 2015), demonstrate poorer emotion regulation in response to distressing material (Davis et al., 2014), and generate fewer problem solutions (Pollock & Williams, 2004). However, the direction of relationships between risk factors and self-injury, and mediators of these relationships, remain unclear. Nonetheless, existing literature indicates that people who self-injure experience various stressful events and are likely to experience maladaptive cognitions, emotion regulation difficulties, and problem-solving deficits. This suggests that psychotherapies which facilitate improvements in these factors may promote coping and thereby reduce the risk of self-injury. Identifying which psychotherapies

effectively improve these outcomes among patients who self-injure could have important ramifications for effectively treating self-injury. Furthermore, investigating which psychotherapies promote change in proposed psychological mechanisms underlying self-injury could improve understanding of mechanisms of change in therapy, thereby allowing therapies to be modified to enhance their effectiveness (Arensman et al., 2001).

Two recent reviews evaluated the impact of psychotherapy on both self-injury and psychological wellbeing outcomes. Hawton et al. (2016) reviewed 29 RCTs of psychosocial interventions for adults with recent self-injury published between 1998 and 2015. Interventions were included where ≥3 RCTs evaluated them. CBT was found to facilitate improvements in problem-solving, depression, and hopelessness. There was little evidence for the effectiveness of DBT on psychological wellbeing outcomes. However, only "one or two" (p. 743) DBT trials were included and details of these trials were not reported. The second review specifically analysed suicidal ideation, depression, and hopelessness outcomes among adults with recent self-injury (Hetrick, Robinson, Spittal, & Carter, 2016). RCTs of psychosocial interventions published between 1999 and 2016 were included. The results suggested treatment effects on each outcome. However, the specific effects of psychotherapies were unclear, as studies of well-recognised therapies (e.g., PST) and less clearly defined psychosocial interventions were conflated. Furthermore, clinical subpopulations were excluded (e.g., patients with borderline personality disorder [BPD]), thus therapies designed specifically for populations who self-injure (e.g., DBT) were not evaluated.

In summary, there is preliminary evidence that psychotherapy facilitates improvement in suicidal ideation, hopelessness, depression, and problem-solving among adults with recent self-injury. However, existing reviews have focused primarily on self-injury reductions rather than detailed investigations of psychological wellbeing

outcomes. It remains unclear which psychotherapies are most beneficial for psychological wellbeing, the breadth of psychological wellbeing outcomes impacted by psychotherapy, and whether the effectiveness of psychotherapy differs between populations presenting with SSI and NSSI.

Current Review

This review aimed to systematically identify and evaluate evidence for the effectiveness of psychotherapies in improving psychological wellbeing among adults with recent self-injury. Building on recent reviews, it aimed to provide a detailed analysis of a wider range of psychological wellbeing outcomes. Psychological wellbeing outcomes were defined as measures of stress, distress, coping, psychological wellbeing, quality of life, and mental health symptoms. This review also aimed to extend findings of recent reviews by including publications from a wider date range (1990-2016), studies of clinical subpopulations, and non-RCT designs. Finally, this review aimed to separately evaluate studies of SSI and NSSI.

Method

The MEDLINE, CINAHL, and PsycINFO databases were searched to identify literature published between the 1st January 1990 and 29th February 2016. Search terms and Boolean operators were (suicid* OR self-harm OR self-injur* OR self-mutilat* OR self-poison* OR parasuicid* OR self-wound*) AND (intervention OR treatment OR *therapy NOT gene therapy). The term "NOT gene therapy" was added after an initial search identified many studies of suicide gene therapy for cancer. Terms were searched for in titles. Forward and backward reference searches were conducted for included studies on the 22nd December 2016.

Definitions of Self-Injury

SSI was defined as an attempt to end one's life for purposes not socially sanctioned (i.e., euthanasia). NSSI was defined as deliberate self-inflicted destruction of

body tissue without suicidal intent (Butler & Malone, 2013). The term non-differentiated self-injury was adopted where studies did not differentiate between SSI and NSSI.

Inclusion and Exclusion Criteria

Included studies (a) were available in English; (b) were published in peer-reviewed journals; (c) used experimental, quasi-experimental, or observational quantitative designs (single case designs needed multiple baseline assessments for reliability purposes); (d) provided data for participants aged ≥16 years; (e) included participants who reported self-injury in the past six months; (f) used standardised self-report measures of stress, distress, coping, psychological wellbeing, quality of life, and/or mental health symptoms pre- and post-therapy; and (g) evaluated community-delivered psychotherapies based on well-defined and recognisable psychological principles. Excluded studies included (a) participants with cognitive impairments or developmental differences (e.g., autism spectrum disorder) and/or (b) only suicidal ideation outcomes.

Procedure

Titles, abstracts, and publication details were reviewed against inclusion and exclusion criteria. Where studies appeared to meet criteria, or there was insufficient information to determine their relevance, full texts were reviewed against inclusion and exclusion criteria.

Quality Appraisal

Included studies were appraised using Downs and Black's (1998) checklist (see Appendix A). This checklist is suitable for appraising randomised and nonrandomised studies and is recommended for use in systematic reviews (Centre for Reviews and Dissemination, 2009). Downs and Black (1998) reported adequate internal consistency (*KR-20* = .89), test-retest reliability (*r* = .88), inter-rater reliability (*r* = .75), and

criterion validity for the checklist. As in previous studies (Larson, Vos, & Fernandez, 2013; Samoocha, Bruinvels, Elbers, Anema, & van der Beek, 2010), item 27 was simplified. Studies scored 1 where power was attained or 0 where power was not attained or assessed. The maximum score was 28. Total scores were labelled to indicate poor quality (0-14 points), fair quality (15-19 points), good quality (20-25 points), and excellent quality (26-28 points). Whilst total scores are crude estimates of quality, they enabled between-study comparisons using qualitative descriptions consistent with previous research (Larson et al., 2013; Samoocha et al., 2010). An independent assessor repeated the quality appraisal on 20% (n = 4) of the included studies. Inter-rater agreement was estimated by calculating a two-way random intraclass correlation coefficient (ICC). Disagreements were resolved through discussion.

Results

Figure 1 depicts the search in a diagram adapted from Liberati et al. (2009). The search identified 3,926 citations. After removing duplicates, 2,392 abstracts were screened and 2,070 were excluded due to irrelevance. Full texts were read for the remaining 322 papers and 24 met criteria for inclusion. Five secondary analyses of included data, which provided no additional relevant information, were also excluded. Of the 19 included studies, 5 had overlapping data with each providing different information.

Across the 16 non-overlapping studies, sample sizes ranged from 18 to 684 (*Mdn* = 56). The total sample size was 2,579. Fifteen studies provided gender ratios, across which participants included 1,795 women and 750 men. Participants were aged between 16 and 66 years. Eight studies reported mean ages ranging from 23.1 to 34.3 years.

The quality, methods, and results of included studies are described below. Studies are ordered by therapeutic modality.

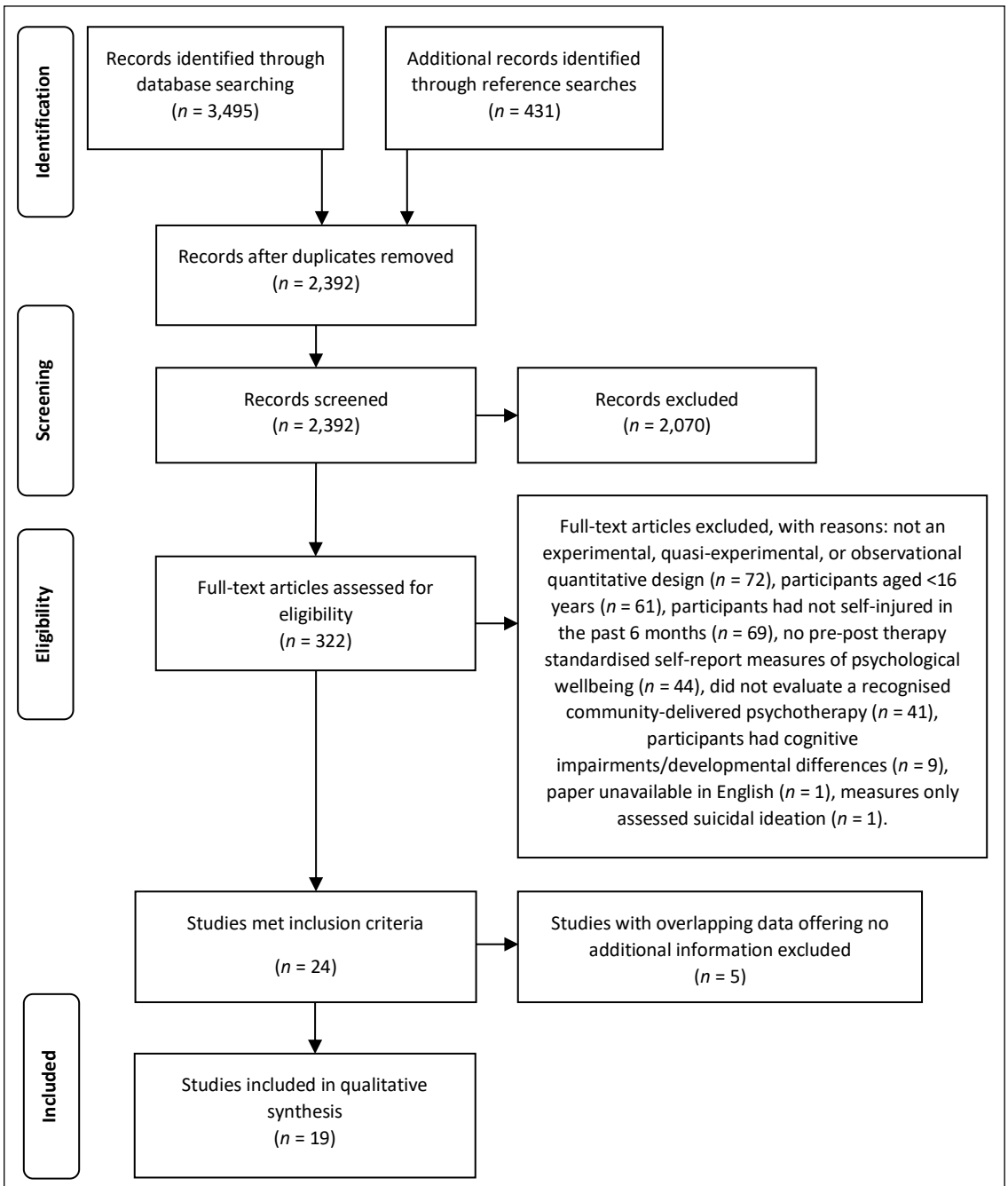

Figure 1. Preferred Reporting Items for Systematic Reviews and Meta-Analyses diagram of the study selection process.

Quality Analysis

Figure 2 presents the quality analysis results. The quality of reporting was generally good. However, only three studies reported monitoring adverse effects by assessing therapy experiences and repeated self-injury. Therefore, iatrogenic effects may not have been fully captured. Also, only five studies described characteristics of participants lost to follow-up. This limited inferences about the generalisability of results and the acceptability of therapies for different populations.

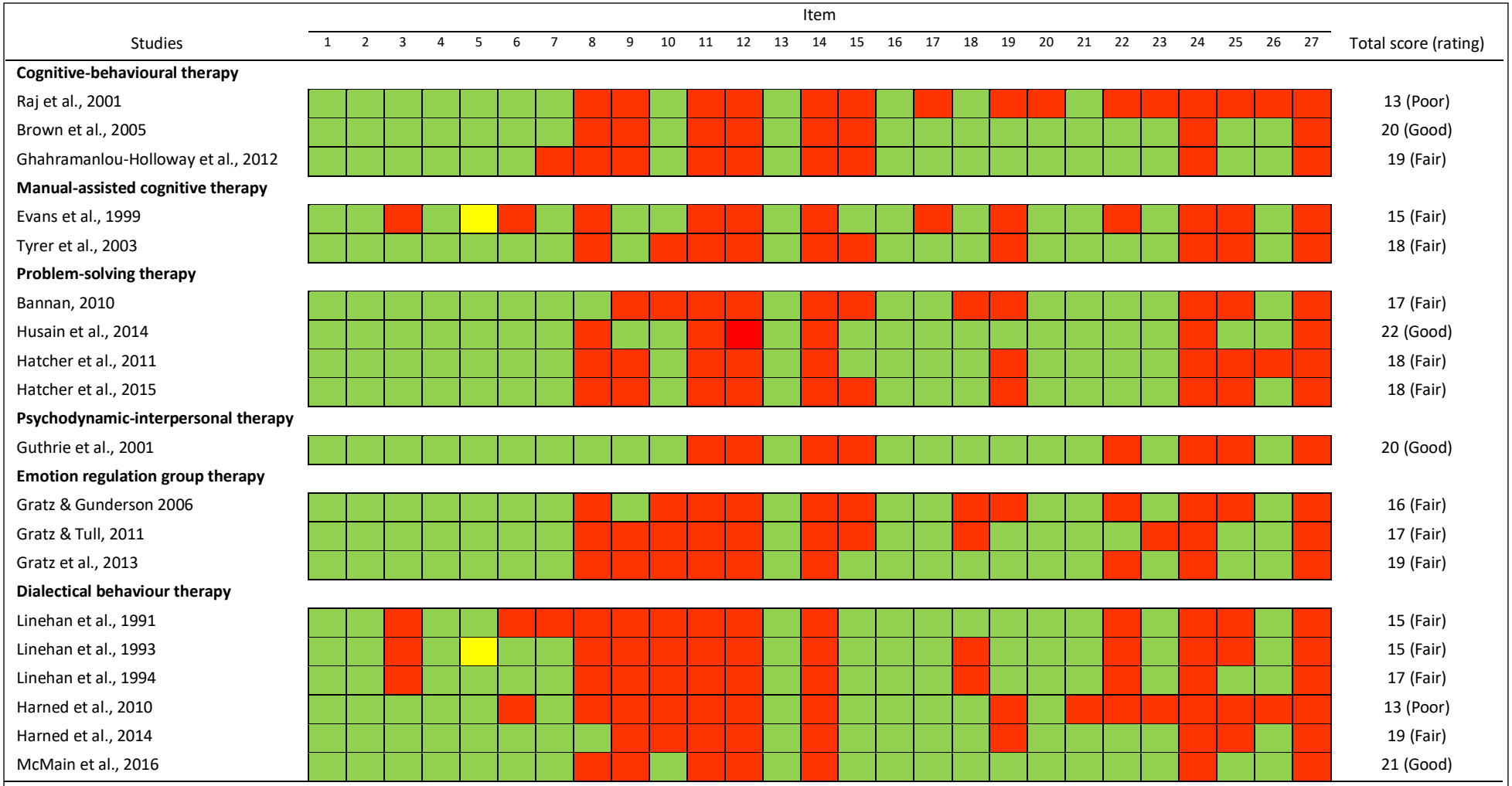

Figure 2. Quality analysis results. ■= Full points. ■ = Partial points. ■ = No points. Items scored 1 where criteria were met except on item 5, where the maximum score was 2. Items scored 0 where criteria were not met or this was undeterminable. Items assessed reporting (1-10), external validity (11-13), bias (14-20), confounding (21-26), and power (27).

The external validity appeared further limited. Whilst all studies evaluated ecologically valid therapies, none recruited participants that were wholly representative of the source population. This limits the extent to which results may be generalised to naturalistic populations.

The risk of bias appeared low. Most studies adjusted for differences in follow-up lengths (n = 17) and reported using validated measures (n = 18). Whilst no studies concealed treatment allocation from participants, ten studies included blind outcome assessors. Furthermore, only self-reported outcomes were reviewed, which limited opportunities for researcher bias.

The risk of confounding was moderate. Sixteen studies were RCTs, which reduced the risk of treatment effects being attributable to confounds, regression to the mean, or temporal change. However, only seven studies reported adjusting for confounding in analyses. This enhanced the risk of treatment effects being attributable to confounds, and may reduce the generalisability of results.

No studies reported sufficient power to detect significant changes on psychological wellbeing measures. However, most studies focused primarily on self-injury reductions and many reported adequate power for this outcome. Subsequently, this quality analysis may underestimate the true power of included studies. This was considered appropriate to avoid overestimating study quality.

Based on the primary author's quality analysis, an independent assessor appraised the quality of one good, one poor, and two fair quality studies. The independent assessor was blind to the primary author's ratings. The ICC was .71, indicating moderate inter-rater agreement (Koo & Li, 2016). Disagreements were resolved through discussion until a consensus on ratings was achieved. Overall, four studies were good quality, thirteen were fair quality, and two were poor quality. Across studies of different psychotherapies, no modality elicited substantially higher

or lower scores (*Mdn*s = 16-20). Quality scores significantly positively correlated with publication date, $r = .52$, $p = .02$.

Methods and Results of Included Studies

The methods and results of included studies are presented in Table 1 and summarised below, categorised by therapeutic modality. Only results on psychological wellbeing outcomes are presented. Table 1 reports whether data were analysed for participants who completed assessments ("completers") or using an intent-to-treat (ITT) paradigm. ITT definitions varied across studies. This review defined ITT as analyses of all data regardless of dropout, compliance, or outcome completion. Significance values, effect sizes, and rates of reliable and clinically significant improvement (RCSI) are reported where available. Studies assessed RCSI using Jacobson and Truax's (1991) criteria. RCSI represents changes in outcome scores from clinical to nonclinical levels whilst taking into account measurement error.

Cognitive-behavioural therapy (CBT). Three studies investigated CBT's impact on psychological wellbeing among participants with SSI. Raj, Kumaraiah, and Bhide (2001) reported significant treatment effects on depression, anxiety, and hopelessness. However, this study received one of the lowest quality scores and its results must be interpreted tentatively. Brown et al.'s (2005) higher quality study supported a treatment effect on depression. However, a marginal treatment effect on hopelessness was not sustained at longer follow-ups. Raj et al. (2001) also reported a treatment effect on problem-solving. This was not supported in Ghahramanlou-Holloway, Bhar, Brown, Olsen, and Beck's (2012) secondary analysis of Brown et al.'s (2005) data, although problem-solving improved significantly faster in the treatment group relative to the control group.

Table 1

Methods and Results of Included Studies

Authors (Country)	Sample	Design	Therapy vs. Comparison	Measures	Results
Cognitive-behavioural therapy (CBT)					
Raj et al., 2001 (India)	40 participants (23 female) aged 16-50 presenting to hospital with suicidal self-injury	Nonrandomised controlled study	CBT vs. TAU	HS DAS PSI HADS	Using completers' data (n = 40), chi-square, Fisher's exact tests and t-tests indicated that the treatment group reported significant pre-post improvements in hopelessness (p = .01), anxiety (p < .001), depression (p = .001), dysfunctional attitudes (p < .001), and problem-solving (p < .001) relative to the control group.
Brown et al., 2005 (US)	120 participants (73 female) aged 18-66 presenting to hospital with suicidal self-injury	RCT	Cognitive therapy vs. enhanced TAU	BDI-II BHS	Using ITT data, hierarchical linear random-effects models indicated that the treatment group reported significant improvements in depressive symptoms 6-months (p = .02) and 12-months (p = .009) post-baseline, relative to the control group. A marginally significant treatment effect emerged on hopelessness 6-months post-baseline (p = .045), but not at later follow-ups.
Ghahramanlou-Holloway et al., 2012 (US)	Secondary analysis of Brown et al.'s (2005) data			SPSI-R-SF	Using ITT data, ANOVAs indicated no significant between-group differences in problem-solving over time. Problem-solving significantly improved temporally in both groups (p < .001). A hierarchical linear random-effects model indicated a faster rate of change in the treatment group, relative to the control group, from baseline to 6-month follow-up in negative problem orientation (p = .02) and impulsive/careless problem-solving (p = .008) scores.
Manual-assisted cognitive therapy (MACT)					
Evans et al., 1999 (United Kingdom)	34 participants aged 16-50 presenting to hospital with self-injury and personality disorder-related difficulties	Pilot RCT	MACT vs. TAU	HADS SFQ	Using completers' data (n = 32), ANCOVAs (adjusted for baseline scores) indicated that the treatment group reported significantly improved depressive symptoms relative to the control group at follow-up (p = .03). Follow-up periods differed between participants. Changes in anxiety and social functioning were nonsignificant.
Tyrer et al., 2003 (United Kingdom)	480 participants (68% female, mean age = 32) presenting to hospital with self-injury	RCT	MACT vs. TAU	HADS HS SFQ EQ-5D	Analyses used completed outcome measurements and sample sizes differed between analyses. Regression analyses indicated no significant between-group differences at 6- or 12-month follow-ups.

(continued)

Authors (Country)	Sample	Design	Therapy vs. Comparison	Measures	Results
Problem-solving therapy (PST)					
Bannan, 2010 (Ireland)	18 women aged 18-53 presenting to hospital after self-poisoning	Pilot RCT	Group PST vs. TAU	BDI HS SPSI-R	Using completers' data (n = 18), ANOVAs indicated that the treatment group reported significant improvements in negative problem orientations relative to the control group post-treatment and at 2-month follow-up. Beneficial treatment effects on depressive symptoms, hopelessness, rational problem-solving, impulsive/careless problem-solving, and avoidant problem-solving were nonsignificant post-treatment but significant at 2-month follow-up.
Husain et al., 2014 (Pakistan)	221 participants (152 female, mean age = 23.1) presenting to hospital with self-injury	RCT	Culturally-adapted PST vs. TAU	BDI BHS EQ-5D CRI	Completers' data were analysed at 3-month (n = 217) and 6-month (n = 213) follow-ups. ANCOVAs (adjusted for baseline scores) indicated that the treatment group reported significant improvements in hopelessness (p = .004), coping (p = .011), and quality of life (p = .012) relative to the control group at 3-month follow-up. At 6-month follow-up, the treatment group reported significant improvements in hopelessness (p = .003), coping (p = .004), quality of life (p = .022), and depressive symptoms (p = .044) relative to the control group.
Hatcher et al., 2011 (New Zealand)	552 participants (380 female, treatment group mean age = 33.2, control group mean age = 34.2) presenting to hospital with self-injury	Zelen RCT	PST vs. TAU	HS HADS SPSI-R	Using completers' data (n = 485), repeated mixed-model regression analyses (adjusted for baseline scores) indicated that the treatment group reported significant improvements on all measures at 3- and 12-month follow-ups (ps ≤ .01). Logistic regression analyses indicated that the proportion of participants with severe hopelessness (HS ≥9) was significantly lower in the treatment group, relative to the control group, at 3-month (odds ratio = 0.24, p < .001) and 12-month (odds ratio = 0.62, p = .03) follow-ups. Relative to participants who attended ≤3 sessions, participants who attended ≥4 sessions had greater improvements in problem-solving (p = .003), anxiety (p = .01), and depression (p = .02) at 3-months, and greater improvements in hopelessness (p = .02) and depression (p = .02) at 12-months.
Hatcher et al., 2015 (New Zealand)	684 participants (464 female, treatment group mean age = 37.5, control group mean age = 36.2), presenting to hospital with self-injury	Zelen RCT	PST+Enhanced Care vs. TAU	HS HADS EQ-5D SF-36	Analyses used completed outcome data and sample sizes differed between analyses. Mixed-model regression analyses indicated no significant between-group differences at 3- or 12-month follow-ups.
Psychodynamic-interpersonal therapy					
Guthrie et al., 2001 (United Kingdom)	119 participants (66 female, mean age = 31.2) presenting to hospital after self-poisoning	RCT	Psychodynamic-interpersonal therapy vs. TAU	BDI	Completers' data were analysed post-treatment (n = 88) and at 6-month follow-up (n = 95). ANCOVAs indicated no significant between-group difference after adjusting for marital status.

(continued)

Authors (Country)	Sample	Design	Therapy vs. Comparison	Measures	Results
Emotion regulation group therapy (ERGT)					
Gratz & Gunderson 2006 (US)	22 self/clinician-referred women (mean age = 33.32) with recent self-injury and BPD-related difficulties	Pilot RCT	ERGT+TAU vs. TAU	DERS AAQ BEST[a] DASS	Using completers' data (n = 22), ANCOVAs (adjusted for baseline scores) indicated that the treatment group reported significant pre-post improvements with large effect sizes in emotion dysregulation ($\acute{\eta}^2_p$ = .54, p < .01), experiential avoidance ($\acute{\eta}^2_p$ = .78, p < .01), BPD symptom severity ($\acute{\eta}^2_p$ = .34, p < .01), depression ($\acute{\eta}^2_p$ = .30, p < .05), anxiety ($\acute{\eta}^2_p$ = .31, p < .01), and stress ($\acute{\eta}^2_p$ = .33, p < .01) relative to the control group. RCSI rates were 83% for emotion dysregulation and experiential avoidance, 50% for depression, 17% for anxiety, and 42% for stress.
Gratz & Tull, 2011 (US)	23 self/clinician-referred women (mean age = 34.3) with recent self-injury and BPD-related difficulties.	Uncontrolled pre-post study	ERGT+TAU	BEST[b] BDI DASS SDS QOLI DERS AAQ	Using completers' data (n = 19), ANOVAs indicated significant pre-post improvements (ps < .05) with large effect sizes in emotion dysregulation ($\acute{\eta}^2_p$ = .67, RCSI = 57.9%), experiential avoidance ($\acute{\eta}^2_p$ = .68, RCSI = 68.4%), BPD symptoms, ($\acute{\eta}^2_p$ = .45, RCSI = 26.3%), BDI-rated depression ($\acute{\eta}^2_p$ = .30, RCSI = 31.6%), DASS-rated depression ($\acute{\eta}^2_p$ = .58, RCSI = 21.1%) anxiety ($\acute{\eta}^2_p$ = .29, RCSI = 10.5%), stress ($\acute{\eta}^2_p$ = .43, RCSI = 21.1%), and social/vocational impairment ($\acute{\eta}^2_p$ = .62, RCSI = 36.8%). Improvements in quality of life ($\acute{\eta}^2_p$ = .10, RCSI = 10.5%) were nonsignificant. Using ITT data, all results were replicated except the treatment effect on anxiety, which became nonsignificant.
Gratz, Tull, & Levy, 2013 (US)	61 self/clinician-referred women (treatment group mean age = 33.3, control group mean age = 33.0) with recent self-injury and BPD-related difficulties	RCT & uncontrolled follow-up	ERGT+TAU vs. TAU	BEST[b] BDI-II DASS IIP-BPD SDS QOLI DERS AAQ	Using ITT data, latent growth models indicated significant (ps < .05) medium-sized treatment effects on emotion dysregulation (d = -0.55), DASS-rated depression (d = -0.51), stress (d = -0.60), and quality of life (d = 0.52). Medium-sized treatment effects for experiential avoidance (d = -0.71) and interpersonal functioning (d = -0.48) were approaching significance. Latent growth models of uncontrolled follow-up data (n = 51) indicated significant improvements (ps <.05) from post-treatment to 9-month follow-up in emotion dysregulation, experiential avoidance, BPD symptoms, and quality of life. Pre-treatment to follow-up effect sizes were medium-large across measures (d = 0.72 – 0.98). RCSI rates improved on all measures from post-treatment (3.2% - 35.5%) to follow-up (13.7% - 49.0%).
Dialectical behaviour therapy (DBT)					
Linehan et al., 1991 (US)	44 clinician-referred women aged 18-45 with recent self-injury and BPD-related difficulties	RCT	DBT vs. TAU	BDI HS	Using completers' data (n = 44), ANCOVAs (adjusted for baseline scores) indicated significant pre-post improvements in depression (p < .005) and hopelessness (p < .05) across the whole sample. No significant between-group differences emerged.

(continued)

Authors (Country)	Sample	Design	Therapy vs. Comparison	Measures	Results
Linehan et al., 1993 (US)	Subset analysis (n = 20) of Linehan et al.'s (1991) data			STAS-T SAS-SR	Analyses used completed outcome measures and sample sizes differed between analyses. ANCOVAs (adjusted for baseline scores) indicated that the treatment group reported significantly better anger ($p < .05$) and social adjustment ($p < .05$) 6-months post-treatment. No significant treatment effects emerged 12-months post-treatment.
Linehan et al., 1994 (US)	Subset analysis (n = 26) of Linehan et al.'s (1991) data			STAS-T SAS-SR	Using ITT data, ANCOVAs (adjusted for baseline scores) indicated that the treatment group reported significantly improved anger ($p < .01$) relative to the control group posttreatment. Improvements on social adjustment were nonsignificant.
Harned et al., 2010 (US)	51 women (mean age = 29.2) with recent self-injury and BPD (26 with comorbid PTSD)	Retrospective case-control analysis of RCT treatment arm data	DBT	DES-T	Analyses used completed outcome measures and sample sizes differed between analyses. Binary hierarchical linear model analyses indicated significant pre-post reductions in severe dissociation (DES-T ≥20) among participants with PTSD ($p < .05$) but not among participants without PTSD.
Harned et al., 2014 (US)	26 self/clinician-referred women (mean age = 32.6) with BPD and PTSD and recent self-injury	Pilot RCT	DBT+Exposure vs. DBT	DES-T TRGI ESS GSI	Using ITT data, mixed-effects models indicated significant temporal improvements in dissociation (p <.05), shame ($p < .001$), and psychological wellbeing ($p < .01$). There was no significant main effect of condition on outcomes. Medium-large pre-post effect sizes were found across measures (gs = 0.4 – 2.1). Between-group effect sizes suggested larger improvements in the DBT+Exposure group, relative to the DBT group, for dissociation, trauma-related guilt cognitions, psychological wellbeing, and shame posttreatment (gs = 0.2 – 1.0), and for dissociation, trauma-related guilt cognitions, and psychological wellbeing at three-month follow-up (gs = 0.2 – 0.7). Posttreatment RCSI rates were higher in the DBT+Exposure group, relative to the DBT group, for trauma-related guilt cognitions (27.3% versus 16.7%), shame (63.6% versus 33.3%), and psychological wellbeing (41.7% versus 0.0%). Follow-up RCSI rates were higher in the DBT+Exposure group, relative to the DBT group, for psychological wellbeing (25.0% versus 0.0%), but were comparable for trauma-related guilt cognitions (18.2% versus 16.7%) and shame (54.5% versus 50.0%).
McMain et al., 2016 (Canada)	84 participants (66 female, mean age = 29.67) with BPD and recent self-injury	RCT	DBT skills training group vs. TAU	BSL-23 BIS-11 SAS-SR STAXI SCL-90-R DERS DTS	Using ITT data, mixed-effects linear growth curve analyses indicated significant treatment effects at 20- and 32-week follow-ups on anger (ps < .001, d = 0.8), emotion dysregulation (ps < .01, d = 0.5), and distress tolerance (ps < .005, d = 0.56). Treatment effects were evident at 20-weeks, but not 32-weeks, on social adjustment ($p < .02$, d = 0.45), symptom distress ($p < .005$, d = 0.41), and BPD symptoms ($p < .01$, d = 0.32). Relative to the control group, the treatment group reported significantly more RCSI in general symptom difficulties at 20-week (43.8% vs. 18.4%, odds ratio = 3.44, p = 0.02), but not at 32-week follow-up.

(continued)

Note. TAU = Treatment-as-usual. US = United States. BPD = Borderline personality disorder. PTSD = Posttraumatic stress disorder. *d* = Cohen's *d*. *g* = Hedge's *g*. $\acute{\eta}^2_p$ = partial eta-squared. ANOVAs = Analyses of variance. ANCOVAs = Analyses of covariance. RCSI = reliable and clinically significant improvement. HS = Hopelessness Scale (Beck, Weissman, Lester, & Trexler, 1974). DAS = Dysfunctional Attitude Scale (Weissman & Beck, 1992). PSI = Problem Solving Inventory (Heppner & Petersen, 1982). HADS = Hospital Anxiety and Depression Scale (Zigmond & Snaith, 1983). BDI-II = Beck Depression Inventory-Second Edition (Beck, Steer, & Brown, 1996). BHS = Beck Hopelessness Scale (Beck & Steer, 1988). SPSI-R-SF = Social Problem-Solving Inventory-Revised-Short-Form (D'Zurilla, Nezu, & Maydeu-Olivares, 2002). SFQ = Social Functioning Questionnaire (Tyrer, 1990). EQ-5D = European Quality of Life (Brooks, 1996; Rabin & de Charro, 2001). BDI = Beck Depression Inventory (Beck, Ward, Mendelsohn, Mock, & Erbaugh, 1961). SPSI-R = Social Problem Solving Inventory-Revised (Chang & D'Zurilla, 1996). CRI = Coping Resource Inventory (Martin & Hammer, 1988). SF-36 = Social Functioning 36-Item Questionnaire (Ware & Sherbourne, 1992). DERS = Difficulties in Emotion Regulation Scale (Gratz & Roemer, 2004). AAQ = Acceptance and Action Questionnaire (Hayes et al., 2004). BEST[a] = Borderline Evaluation of Severity over Time (Pfohl & Blum, 1997). BEST[b] = Borderline Evaluation of Severity over Time (Pfohl et al., 2009). DASS = Depression Anxiety Stress Scales (Lovibond & Lovibond, 1995). SDS = Sheehan Disability Scale (Sheehan, 1983). QOLI = Quality of Life Inventory (Frisch, Cornwell, Villanueva, & Retzlaff, 1992). IIP-BBD = Inventory of Interpersonal Problems-Borderline Composite (Lejuez et al., 2003). STAS-T = State-Trait Anger Scale (Spielberger, Jacobs, Russell, & Crane, 1983). SAS-SR = Social Adjustment Scale-Self-Report (Weissman & Bothwell, 1976). DES-T = Dissociative Experiences Scale–Taxon (Waller & Ross, 1997). TRGI = Trauma-Related Guilt Inventory (Kubany et al., 1996). ESS = Experience of Shame Scale (Andrews, Qian, & Valentine, 2002). GSI = Brief Symptom Inventory-General Psychological Wellbeing subscale (Derogatis, 1993). BSL-23 = Borderline Symptom List-23 (Bohus et al., 2009). BIS-11 = Barrett Impulsiveness Scale-11 (Patton, Stanford, & Barratt, 1995). STAXI = State-Trait Anger Expression Inventory (Spielberger, 1988). SCL-90-R = Symptom Checklist-90-Revised (Derogatis & Melisaratos, 1983). DTS = Distress Tolerance Scale (Simons & Gaher, 2005).

Manual-assisted cognitive therapy (MACT). Two studies investigated MACT's impact on psychological wellbeing among participants with non-differentiated self-injury. Whilst a pilot trial reported a significant treatment effect on depression (Evans et al., 1999), this was not replicated in a more robust RCT of the intervention (Tyrer et al., 2003). Neither study found treatment effects on anxiety or social functioning.

Problem-solving therapy (PST). Four studies investigated PST's impact on psychological wellbeing among participants with non-differentiated self-injury. Across three studies, findings indicated treatment effects on mental health symptoms, problem-solving, and quality of life (Bannan, 2010; Hatcher, Sharon, Parag, & Collins, 2011; Husain et al., 2014). Treatment effects emerged at follow-up rather than post-treatment in two studies, particularly on measures of depression (Bannan, 2010; Husain et al., 2014). However, the largest RCT found no treatment effects on depression, hopelessness, anxiety, or quality of life (Hatcher et al., 2015). Nonetheless, Husain et al.'s (2014) study had the highest quality score, indicating that its results may be more valid and reliable relative to the other studies.

Psychodynamic-interpersonal psychotherapy. One study investigated psychodynamic-interpersonal psychotherapy's impact on depression among participants with non-differentiated self-injury. No treatment effect emerged after adjusting for marital status (Guthrie et al., 2001).

Emotion regulation group therapy (ERGT). Three studies investigated ERGT's impact on psychological wellbeing among participants with non-differentiated self-injury. Two studies reported large treatment effects on emotion dysregulation, experiential avoidance, depression, stress, and BPD symptoms (Gratz & Gunderson, 2006; Gratz & Tull, 2011). Furthermore, ≥50% of participants reported RCSI on hypothesised mechanisms of change (i.e., emotion dysregulation and experiential

avoidance). These studies were small and only Gratz and Gunderson (2006) included a control group. The results were partially supported by a larger and higher quality study, which found medium-sized treatment effects on emotion dysregulation, depression, stress, and quality of life (Gratz, Tull, & Levy, 2013). Treatment effects were not replicated for experiential avoidance or BPD symptoms. RCSI rates across measures increased over follow-up, suggesting a strengthening of treatment effects.

Dialectical behaviour therapy (DBT). Six studies evaluated DBT's impact on psychological wellbeing among participants with non-differentiated self-injury. Across three overlapping studies (Linehan, Armstrong, Suarez, Allmon, & Heard, 1991; Linehan, Heard, & Armstrong, 1993; Linehan, Tutek, Heard, & Armstrong, 1994), treatment effects emerged on anger and social adjustment 6-months post-treatment. These effects were not maintained 12-months post-treatment. No treatment effects emerged on depression or hopelessness post-treatment, and no follow-up data were reported for these variables. These results were partially supported in a more diverse sample who received DBT group skills training (McMain, Guimond, Barnhart, Habinski, & Streiner, 2016). Treatment effects were sustained at 3-month follow-up for anger, distress tolerance, and emotion regulation, but not for social adjustment, symptom distress, or BPD symptoms.

The two remaining studies included women with posttraumatic stress disorder (PTSD) and BPD. Harned, Jackson, Comtois, and Linehan (2010) found a treatment effect on severe dissociation among participants with comorbid BPD and PTSD, but not among participants with non-comorbid BPD. However, participants with PTSD reported greater dissociation pretherapy. Consequently, regression to the mean may have impacted results. Nonetheless, Harned, Korslund, and Linehan (2014) also found treatment effects on dissociation, shame, and psychological wellbeing among patients with comorbid BPD and PTSD, although effect sizes and recovery rates were larger for

participants who received additional exposure interventions.

Discussion

This review aimed to systematically identify and evaluate studies investigating the impact of psychotherapies on psychological wellbeing among adults with recent self-injury. The search identified 2,392 unique papers, of which 19 were reviewed. This review initially aimed to evaluate studies of SSI and NSSI separately. However, only three studies distinguished these presentations, and all three focused on SSI. Given that SSI and NSSI may represent distinct needs (Nock, 2009), future studies should differentiate these presentations and investigate differences in their treatment responses. The results of reviewed studies are discussed below, categorised by therapeutic modality.

Cognitive-Behavioural Therapy (CBT)

All three studies of CBT included adults with recent SSI. Findings suggested a treatment effect on depression (Brown et al., 2005; Raj et al., 2011), which supports previous findings (Hawton et al., 2016). Improvements in other mental health symptoms (e.g., hopelessness) were not sustained long-term or were limited to a poor-quality study. In contrast, Hawton et al.'s (2016) previous review indicated a treatment effect on hopelessness. Inspection of the studies Hawton et al. (2016) categorised as CBT revealed studies of psychodynamic-interpersonal therapy, PST, and integrative therapy. Two studies of traditional CBT were included which did not meet the current review's inclusion criteria (Stewart, Quin, Plever, & Emmerson, 2009; Wei et al., 2013). Nonetheless, these studies also found no treatment effect on hopelessness or quality of life among adults with SSI. Therefore, the treatment effect on hopelessness in Hawton et al.'s (2016) review may have been attributable to non-traditional CBT modalities.

Findings regarding CBT's effect on problem-solving were inconsistent. Raj et al. (2001) found a significant treatment effect whereas Ghahramanlou-Holloway et al.

(2012) did not. Again, whilst Hawton et al. (2016) reported a treatment effect for CBT on problem-solving, this appeared attributable to psychotherapies that were examined separately in the current review (e.g., PST). There are several hypotheses for the current review's inconsistent findings. Raj et al. (2001) used a nonrandomised design and the treatment group reported poorer problem-solving than the control group pretherapy. As this difference was not adjusted for, results may have been affected by regression to the mean. Also, the comparison conditions that studies employed may have influenced results. The comparison conditions were routine medical care (Raj et al., 2001) or case management (Ghahramanlou-Holloway et al., 2012). A previous study of participants with depression found that psychotherapy and case management facilitated comparable improvements in problem-solving (Areán et al., 2015). Therefore, Ghahramanlou-Holloway et al.'s (2012) participants may have received beneficial interventions for problem-solving in either condition, thereby accounting for the null treatment effect.

Manual-Assisted Cognitive Therapy (MACT)

Two studies investigated MACT's effectiveness among participants with non-differentiated self-injury. Findings indicated no treatment effects on anxiety or social functioning. Whilst a treatment effect on depression emerged in a pilot trial (Evans et al., 1999), this was not replicated in a more robust trial (Tyrer et al., 2003). As more of Evans et al.'s (1999) participants received direct therapist contact than Tyrer et al.'s (2003) participants, therapist contact may have been particularly beneficial. Indeed, a previous RCT found no treatment effect for online self-help on depression among adults with suicidal thoughts relative to a waitlist control (van Spijker, van Straten, & Kerkhof, 2014). Alternatively, the different findings may relate to the different comparison conditions that studies employed. Comparison groups received referrals for either psychiatric care and psychoeducation (Evans et al., 1999) or alternative therapy (Tyrer et al., 2003). A previous study which referred a comparison group for psychiatric care

also found a treatment effect for MACT on depression (Davidson, Brown, James, Kirk, & Richardson, 2014). Tyrer et al.'s (2003) null effect may therefore have related to participants receiving beneficial therapies regardless of allocation. Accordingly, many participants in the control group received more psychotherapy than participants who received MACT, and descriptive statistics suggested improvements in depressive symptoms in both groups. Within-group analyses of pre-post scores could have confirmed this. Nonetheless, the results suggested little evidence for the effectiveness of MACT above other psychotherapies in improving mental health among adults with non-differentiated self-injury.

Problem-Solving Therapy (PST)

Findings suggested treatment effects for PST on mental health symptoms, quality of life, and its hypothesised mechanism of change, problem-solving (Bannan, 2010; Hatcher et al., 2011; Husain et al., 2014). This supports previous findings (Townsend et al., 2001). There was some evidence that treatment effects may be delayed until follow-up. As PST teaches problem-solving skills, it may be that its effect strengthens temporally as participants master skills. Indeed, a previous meta-analysis found a slightly larger effect size for PST on depression at follow-up compared to posttreatment (Bell & D'Zurilla, 2009).

However, the largest study of PST found no treatment effects (Hatcher et al., 2015). This result may have related to poor treatment compliance. Hatcher et al.'s (2015) participants attended four sessions on average, compared to five (Husain et al., 2014) and six (Hatcher et al., 2011) in other studies. Bannan (2010) did not report attendance rates but described good treatment adherence. Indeed, therapy attendance positively related to outcome improvements in two studies (Hatcher et al., 2011, 2015). Hatcher et al.'s (2015) study also differed from the other studies as it included a comprehensive care package alongside PST. This design prevented the unique effects of

PST to be disentangled. It is possible that one of the additional interventions detrimentally affected outcomes. Investigations of adverse outcomes were rare amongst included studies, and should be measured in future studies to capture iatrogenic effects.

Psychodynamic-Interpersonal Psychotherapy

One study investigated psychodynamic-interpersonal psychotherapy and found no treatment effect on depression (Guthrie et al., 2001). The intervention in this study was brief (i.e., four sessions) and participants presented with moderate-severe depression, which has been found to respond significantly better to longer durations of psychodynamic-interpersonal therapy (Barkham et al., 1996). Accordingly, in secondary analyses of their data, Guthrie et al. (2003) found that the intervention was significantly more effective for participants with less depressive symptoms pre-therapy. Replications of Guthrie et al.'s (2001) study with longer psychotherapy durations could determine whether the null result related to the intervention's brevity.

Emotion Regulation Group Therapy (ERGT)

Findings consistently suggested medium-large treatment effects for ERGT on emotion dysregulation, depression, and stress (Gratz & Gunderson, 2006; Gratz & Tull, 2011; Gratz et al., 2013). RCSI rates across included measures ranged from 3.2% to 83.0%, and there was preliminary evidence that these rates increased temporally. ERGT aims to facilitate change by improving participants' emotion regulation skills. Accordingly, studies support that improvements in emotion regulation mediate outcomes in ERGT (Gratz, Bardeen, Levy, Dixon-Gordon, & Tull, 2015) and are associated with better psychological wellbeing across various clinical presentations (Smyth & Arigo, 2009). However, as all three studies of ERGT included women with BPD-related difficulties, further studies need to ascertain the effectiveness of ERGT with other populations.

Dialectical Behaviour Therapy (DBT)

Treatment effects were reported for DBT on anger, social adjustment, shame, coping, distress, and mental health symptoms (Harned et al., 2010, 2015; Linehan et al., 1993, 1994; McMain et al., 2016). However, comparisons between studies were complicated by their diverse methods, interventions, and samples. The outcome measures varied widely, thereby hampering the ability to ascertain consistent findings. Indeed, previous meta-analyses which have reported null effects for DBT on psychological wellbeing outcomes have encountered heterogeneous studies with inconsistent findings (Hawton et al., 2016; Panos, Jackson, Hasan, & Panos, 2014). Future studies should aim to replicate previously found treatment effects.

There was some evidence that DBT's effect weakened temporally (Linehan et al., 1993; McMain et al., 2016). Whilst previous studies of DBT have indicated sustained treatment effects on self-injury (Gibson, Booth, Davenport, Keogh, & Owens, 2014; Linehan et al., 2006), few studies have reported follow-up data for psychological wellbeing outcomes. In Neacsiu, Rizvi, and Linehan's (2010) study of 108 women, DBT skill use across treatment and a four-month follow-up fully mediated improvements in SSI, depression, and anger, and partially mediated improvements in NSSI. Future studies should measure how skill use relates to outcomes, and evaluate longevity in psychological wellbeing improvements following DBT. Furthermore, as every DBT study included participants with personality disorder-related difficulties and only one included men, future studies could ascertain the effectiveness of DBT for different populations who self-injure.

Critique

A strength of the included studies was that 16 were RCTs (84.21%). This enhanced the ability to infer treatment effects that were not attributable to confounds, regression to the mean, and temporal change. Furthermore, no studies restricted comparison groups from seeking alternative interventions. This may have reduced

between-group differences regarding non-specific therapy factors (e.g., time with therapists), thereby enhancing the ability to attribute effects to the experimental psychotherapies. Also, the results may represent realistic estimates of the experimental psychotherapies' effectiveness relative to standard practice across the populations sampled.

However, only two studies were conducted in non-Western countries (i.e., India and Pakistan). Therefore, the cross-cultural validity of the results is unclear. The generalisability and ecological validity of the results may be further limited by the strict inclusion criteria that studies employed. Whilst controlled trials make important contributions to the literature, more practice-based evaluations of psychotherapy's effectiveness in routine practice are needed. Many studies featured similar inclusion criteria. Typically, studies included treatment-seeking adults who had self-injured in the past six months, were community-dwelling, and did not have schizophrenia, substance misuse difficulties, cognitive impairments, or developmental differences. The results may therefore be representative of this population. However, women were overrepresented. Future studies need to determine psychotherapy's impact on psychological wellbeing among men, particularly for ERGT and DBT. Regarding analyses, effect sizes and recovery rates were inconsistently reported, and mechanisms of change were rarely measured. This hampered the ability to infer the size, and causes, of change. Finally, no studies reported adequate power for detecting significant changes on psychological wellbeing measures, which enhanced the risk of type two error.

Regarding the current review, strengths included that it was undertaken systematically without predetermined hypotheses. Its focus on well-established psychotherapies may facilitate the dissemination and application of its results to practice. Furthermore, heterogeneous studies were included, to provide a comprehensive review of all available studies. Indeed, the database search identified 3,495 citations and

reference searches were also conducted. This reduced the risk of missing relevant papers.

However, as only one person conducted the search, its comprehensive nature enhanced the risk of human error in determining study relevance. Future reviews should include multiple assessors to enhance the reliability of searches. Furthermore, only studies published in peer-reviewed journals were included, which can lead to overestimations of treatment effects (Cuijpers, Smit, Bohlmeijer, Hollon, & Andersson, 2010). Future reviews should include grey literature to reduce the risk of publication bias. Finally, only papers available in English were included, which increased the risk of missing relevant papers.

Conclusions and Implications

This systematic review suggests that psychotherapy facilitates improvements in psychological wellbeing among adults with self-injury. This has implications for the continued provision of psychotherapy for this population. Specifically, the most consistent support was found for the effectiveness of PST and ERGT in improving mental health outcomes, as well as their proposed mechanisms of change. This supports theoretical models proposing the roles of problem-solving (Rudd, 2006; Williams & Pollock, 2001) and emotion regulation (Gratz & Roemer, 2004; Linehan, 1993) in mediating adults' self-injury. Empirical evaluations of the role of these variables in predicting therapy outcomes could support this conclusion. In clinical practice, assessments of problem-solving and emotion regulation could identify treatment needs among patients who self-injure and provide direction towards the most appropriate psychotherapy. Clinicians are also advised to consider ways to enhance treatment compliance, given the positive relationship between PST attendance and outcomes. Furthermore, some treatment effects strengthened temporally or emerged only at follow-up. Consequently, for accurate evaluations of therapeutic change, clinicians are

advised to monitor patient outcomes post-therapy.

Regarding other psychotherapies, CBT appeared effective for improving depression among adults with recent SSI. However, treatment effects on other outcomes were limited. DBT was found to facilitate improvements in psychological wellbeing among predominantly female participants with personality disorder-related difficulties. However, future studies need to replicate, and evaluate the longevity of, treatment effects. Little evidence was found for the superiority of MACT or psychodynamic-interpersonal therapy above other interventions in improving psychological wellbeing. However, this tentative conclusion was based on only three studies of brief interventions, and compliance with MACT was low. Research investigating dose-response relationships between these psychotherapies and psychological wellbeing outcomes could ascertain whether the interventions reviewed were too brief to be effective.

As most studies were RCTs with strict inclusion criteria, more practice-based studies are needed to evaluate the effectiveness of psychotherapy for adults who self-injure in routine practice. The literature would also benefit from more studies with men and studies which differentiate between SSI and NSSI. This is crucial to identify effective therapies for different populations who self-injure. Finally, the current review's results could be extended with an individual patient data meta-analysis, to provide a reliable estimate of the overall effect of psychotherapy on psychological wellbeing among patients who self-injure.

References

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Washington D.C., WA: American Psychiatric Association.

Andrews, B., Qian, M. Y., & Valentine, J. D. (2002). Predicting depressive symptoms with a new measure of shame: The Experience of Shame Scale. *British Journal of Clinical Psychology, 41,* 29-42. doi:10.1348/014466502163778

Areán, P. A., Raue, P. J., McCulloch, C., Kanellopoulos, D., Seirup, J. K., Banerjee, S., … Alexopoulos, G. S. (2015). Effects of problem-solving therapy and clinical case management on disability in low-income older adults. *The American Journal of Geriatric Psychiatry, 23,* 1307-1314. doi:10.1016/j.jagp.2015.04.005

Arensman, E., Townsend, E., Hawton, K., Bremner, S., Feldman, E., Goldney, R., … Träskman-Bendz, L. (2001). Psychosocial and pharmacological treatment of patients following deliberate self-harm: The methodological issues involved in evaluating effectiveness. *Suicide and Life Threatening Behavior, 31,* 169-180. doi:10.1521/suli.31.2.169.21516

Bannan, N. (2010). Group-based problem-solving therapy in self-poisoning females: A pilot study. *Counselling and Psychotherapy Research: Linking Research with Practice, 10,* 201-213. doi:10.1080/14733140903337292

Barkham, M., Rees, A., Stiles, W., Shapiro, D. A., Hardy, G. E., & Reynolds, S. (1996). Dose-effect relations in time-limited psychotherapies for depression. *Journal of Consulting and Clinical Psychology, 64,* 927-935. doi:10.1037//0022-006x.64.5.927

Barzilay-Levkowitz, S., & Apter, A. (2014). Psychological models of suicide. *Archives of Suicide Research, 18,* 295-312. doi:10.1080/13811118.2013.824825

Beck, A. T., Brown, G., Berchick, R. J., Stewart, B. L., & Steer, R. A. (1990). Relationship between hopelessness and ultimate suicide: A replication with

psychiatric outpatients. *American Journal of Psychiatry, 147,* 190-195. doi:10.1176/ajp.147.2.190

Beck, A. T., & Steer, R. A. (1988). *Manual for the Beck Hopelessness Scale*. San Antonio, TX: Psychological Corporation.

Beck, A. T., Steer, R. A., & Brown, G. K. (1996). *Manual for the Beck Depression Inventory II.* San Antonio, TX: Psychological Corporation.

Beck, A. T., Ward, C. H., Mendelsohn, M., Mock, J., & Erbaugh, J. (1961). An inventory for measuring depression. *Archives of General Psychiatry, 4,* 561–71. doi:10.1001/archpsyc.1961.01710120031004

Beck, A. T., Weissman, A., Lester, D., & Trexler, L. (1974). The measurement of pessimism: The Hopelessness Scale. *Journal of Consulting and Clinical Psychology, 42,* 861-865. doi:10.1037/h0037562

Bell, A. C., & D'Zurilla, T. J. (2009). Problem-solving therapy for depression: A meta-analysis. *Clinical Psychology Review, 29,* 348-353. doi:10.1016/j.cpr.2009.02.003

Bohus, M., Kleindienst, N., Limberger, M. F., Stieglitz, R. D., Domsalla, M., Chapman, A. L., … Wolf, M. (2009). The short version of the Borderline Symptom List (BSL-23): Development and initial data on psychometric properties. *Psychopathology, 42,* 32–39. doi:10.1159/000173701

Briere, J., & Gil, E. (1998). Self-mutilation in clinical and general population samples: Prevalence, correlates, and functions. *American Journal of Orthopsychiatry, 68,* 609-620. doi:10.1037/h0080369

Brooks, R. (1996). EuroQol: The current state of play. *Health Policy, 37,* 53–72. doi:10.1016/0168-8510(96)00822-6

Brown, G. K., Have, T. T., Henriques, G. R., Xie, S. X., Hollander, J. E., & Beck, A. T. (2005). Cognitive therapy for the prevention of suicide attempts. A randomized

controlled trial. *Journal of the American Medical Association, 294,* 563-570. doi:10.1001/jama.294.5.563

Butler, A. M., & Malone, K. (2013). Attempted suicide v. non-suicidal self-injury: Behaviour, syndrome or diagnosis? *British Journal of Psychiatry, 202,* 324-325. doi:10.1192/bjp.bp.112.113506

Carroll, R., Metcalfe, C., & Gunnell, D. (2014). Hospital presenting self-harm and risk of fatal and non-fatal repetition: Systematic review and meta-analysis. *PLoS ONE, 9*(2). Retrieved from http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0089944

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

Chang, E. C., & D'Zurilla, T. J. (1996). Relations between problem orientation and optimism, pessimism, and trait affectivity: A construct validation study. *Behaviour Research Therapy, 34,* 185-194. doi:10.1016/0005-7967(95)00046-1

Comtois, K. A., & Linehan, M. M. (2006). Psychosocial treatments of suicidal behaviors: A practice-friendly review. *Journal of Clinical Psychology, 62,* 161-170. doi:10.1002/jclp.20220

Crawford, M. J., Thomas, O., Khan, N., & Kulinskaya, E. (2007). Psychosocial interventions following self-harm. Systematic review of their efficacy in preventing suicide. *British Journal of Psychiatry, 190,* 11-17. doi:10.1192/bjp.bp.106.025437

Cuijpers, P., Smit, F., Bohlmeijer, E., Hollon, S. D., & Andersson, G. (2010). Efficacy of cognitive-behavioural therapy and other psychological treatments for adult depression: Meta-analytic study of publication bias. *British Journal of Psychiatry, 196,* 173-178. doi:10.1192/bjp.bp.109.066001

Daigle, M. S., Pouliot, L., Chagnon, F., Greenfield, B., & Mishara, B. (2011). Suicide

attempts: Prevention of repetition. *Canadian Journal of Psychiatry, 56,* 621-629. Retrieved from http://journals.sagepub.com/doi/pdf/10.1177/070674371105601008

Davidson, K. M., Brown, T. M., James, V., Kirk, J., & Richardson, J. (2014). Manual-assisted cognitive therapy for self-harm in personality disorder and substance misuse: A feasibility trial. *Psychiatric Bulletin, 38,* 108-111. doi:10.1192/pb.bp.113.043109

Davis, T. S., Mauss, I. B., Lumian, D., Troy, A. S., Shallcross, A. J., Zarolia, P., … McRae, K. (2014). Emotional reactivity and emotion regulation among adults with a history of self-harm: Laboratory self-report and functional MRI evidence. *Journal of Abnormal Psychology, 123,* 499-509. doi:10.1037/a0036962

Department of Health. (2012). *Preventing suicide in England: A cross-government outcomes strategy to save lives.* London, UK: Author.

Derogatis, L. R. (1993). *Brief Symptom Inventory: Administration, scoring, and procedures manual.* Bloomington, MN: Pearson.

Derogatis, L. R., & Melisaratos, N. (1983). The Brief Symptom Inventory: An introductory report. *Psychological Medicine, 13,* 595–605. doi:10.1017/s0033291700048017

Downs, S. H., & Black, N. (1998). The feasibility of creating a checklist for the assessment of the methodological quality both of randomised and non-randomised studies of health care interventions. *Journal of Epidemiology & Community Health, 52,* 377-384. doi:10.1136/jech.52.6.377

D'Zurilla, T. J., Nezu, A. M., & Maydeu-Olivares, A. (2002). *Social Problem-Solving Inventory-Revised: Technical manual.* New York, NY: Multi-Health Systems.

Evans, K., Tyrer, P., Catalan, J., Schmit, U., Davidson, K., Dent, J., … Thompson, S. (1999). Manual-assisted cognitive-behaviour therapy (MACT): A randomized

controlled trial of a brief intervention with bibliotherapy in the treatment of recurrent deliberate self-harm. *Psychological Medicine, 29,* 19-25. doi:10.1017/s003329179800765x

Frisch, M. B., Cornwell, J., Villanueva, M., & Retzlaff, P. J. (1992). Clinical validation of the Quality of Life Inventory: A measure of life satisfaction of use in treatment planning and outcome assessment. *Psychological Assessment, 4,* 92–101. doi:10.1037//1040-3590.4.1.92

Ghahramanlou-Holloway, M., Bhar, S. S., Brown, G., Olsen, C., & Beck, A. T. (2012). Changes in problem-solving appraisal after cognitive therapy for the prevention of suicide. *Psychological Medicine, 42,* 1185-1193. doi:10.1017/S0033291711002169

Gibson, J., Booth, R., Davenport, J., Keogh, K., & Owens, T. (2014). Dialectical behaviour therapy-informed skills training for deliberate self-harm: A controlled trial with 3-month follow-up data. *Behaviour Research and Therapy, 60,* 8-14. doi:10.1016/j.brat.2014.06.007

Gratz, K. L., Bardeen, J. R., Levy, R., Dixon-Gordon, K. L., & Tull, M. T. (2015). Mechanisms of change in an emotion regulation group therapy for deliberate self-harm among women with borderline personality disorder. *Behaviour Research and Therapy, 65,* 29-35. doi:10.1016/j.brat.2014.12.005

Gratz, K. L., & Gunderson, J. G. (2006). Preliminary data on an acceptance-based emotion regulation group intervention for deliberate self-harm among women with borderline personality disorder. *Behavior Therapy, 37,* 25-35. doi:10.1016/j.beth.2005.03.002

Gratz, K. L., & Roemer, L. (2004). Multidimensional assessment of emotion regulation and dysregulation: Development, factor structure, and initial validation of the Difficulties in Emotion Regulation Scale. *Journal of Psychopathology and*

Behavioral Assessment, 26, 41–54. doi:10.1023/b:joba.0000007455.08539.94

Gratz, K. L., & Tull, M. T. (2011). Extending research on the utility of an adjunctive emotion regulation group therapy for deliberate self-harm among women with borderline personality pathology. *Personality Disorders: Theory, Research, and Treatment, 2,* 316-326. doi:10.1037/a0022144

Gratz, K. L., Tull, M. T., & Levy, R. (2013). Randomized controlled trial and uncontrolled 9-month follow-up of an adjunctive emotion regulation group therapy for deliberate self-harm among women with borderline personality disorder. *Psychological Medicine, 44,* 2099-2112. doi:10.1017/S0033291713002134

Guthrie, E., Kapur, N., Mackway-Jones, K., Chew-Graham, C., Moorey, J., Mendel, E., … Boddy, G. (2003). Predictors of outcome following brief psychodynamic-interpersonal therapy for deliberate self poisoning. *Australian and New Zealand Journal of Psychiatry, 37,* 532-536. doi:10.1046/j.1440-1614.2003.01197.x

Guthrie, E., Kapur, N., Mackway-Jones, K., Chew-Graham, C., Moorey, J., Mendel, E., … Tomenson, B. (2001). Randomised controlled trial of brief psychological intervention after deliberate self-poisoning. *British Medical Journal, 323,* 1-5. doi:10.1136/bmj.323.7305.135

Harned, M. S., Jackson, S. C., Comtois, K. A., & Linehan, M. M. (2010). Dialectical behaviour therapy as a precursor to PTSD treatment for suicidal and/or self-injuring women with borderline personality disorder. *Journal of Traumatic Stress, 23,* 421-429. doi:10.1002/jts.20553

Harned, M. S., Korslund, K. R., & Linehan, M. M. (2014). A pilot randomized controlled trial of dialectical behaviour therapy with and without the dialectical behaviour therapy prolonged exposure protocol for suicidal and self-injuring women with borderline personality disorder and PTSD. *Behaviour Research and*

Therapy, 55, 7-17. doi:10.1016/j.brat.2014.01.008

Hatcher, S., Sharon, C., House, A., Collins, N., Collings, S., & Pillai, A. (2015). The ACCESS study: Zelen randomised controlled trial of a package of care for people presenting to hospital after self-harm. *British Journal of Psychiatry, 206,* 229-236. doi:10.1192/bjp.bp.113.135780

Hatcher, S., Sharon, C., Parag, V., & Collins, N. (2011). Problem-solving therapy for people who present to hospital with self-harm: Zelen randomised controlled trial. *British Journal of Psychiatry, 199,* 310-316. doi:10.1192/bjp.bp.110.090126

Hawton, K., Bergen, H., Cooper, J., Turnbull, P., Waters, K., Ness, J., & Kapur, N. (2015). Suicide following self-harm: Findings from the multicentre study of self-harm in England, 2000-2012. *Journal of Affective Disorders, 175,* 147-151. doi:10.1016/j.jad.2014.12.062

Hawton, K., Townsend, E., Arensman, E., Gunnell, D., Hazell, P., House, A., & van Heeringen, K. (1999). Psychosocial and pharmacological treatments for deliberate self-harm. *Cochrane Database of Systematic Reviews, 1999*(4), 1-42. doi:10.1002/14651858.cd001764.pub2

Hawton, K., Witt, K. G., Taylor Salisbury, T. L., Arensman, E., Gunnell, D., Hazell, P., … van Heeringen, K. (2016). Psychosocial interventions following self-harm in adults: A systematic review and meta-analysis. *Lancet Psychiatry, 3,* 740-750. doi:10.1016/s2215-0366(16)30070-0

Hayes, S. C., Strosahl, K., Wilson, K. G., Bissett, R. T., Pistorello, J., Toarmino, D., . . . McCurry, S. M. (2004). Measuring experiential avoidance: A preliminary test of a working model. *The Psychological Record, 54,* 553–578. Retrieved from http://opensiuc.lib.siu.edu/cgi/viewcontent.cgi?article=1359&context=tpr

Heppner, P. P., & Petersen, C. H. (1982). The development and implications of a personal problem-solving inventory. *Journal of Counseling Psychology, 29,* 66-

75. doi:10.1037//0022-0167.29.1.66

Hetrick, S. E., Robinson, J., Spittal, M. J., & Carter, G. (2016). Effective psychological and psychosocial approaches to reduce repetition of self-harm: A systematic review, meta-analysis and meta-regression. *British Medical Journal Open, 6.* Retrieved from http://bmjopen.bmj.com/content/bmjopen/6/9/e011024.full.pdf

Husain, N., Afsar, S., Ara, J., Fayyaz, H., ur Rahman, R., Tomenson, B., … Chaudhry, I. B. (2014). Brief psychological intervention after self-harm: Randomised controlled trial from Pakistan. *British Journal of Psychiatry, 204,* 462-470. doi:10.1192/bjp.bp.113.138370

Jacobson, C. M., Muehlenkamp, J. J., Miller, A. L., & Turner, J. B. (2008). Psychiatric impairment among adolescents engaging in different types of deliberate self-harm. *Journal of Clinical Child & Adolescent Psychology, 37,* 363-375. doi:10.1080/15374410801955771

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

Klonsky, E. D. (2011). Non-suicidal self-injury in United States adults: Prevalence, sociodemographics, topography and functions. *Psychological Medicine, 41,* 1981-1986. doi:10.1017/S0033291710002497

Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine, 15,* 155-163. doi:10.1016/j.jcm.2016.02.012

Kubany, E. S., Haynes, S. N., Abueg, F. R., Manke, F. P., Brennan, J. M., & Stahura, C. (1996). Development and validation of the Trauma-related Guilt Inventory (TRGI). *Psychological Assessment, 8,* 428-444. doi:10.1037//1040-3590.8.4.428

Larkin, C., di Blasi, Z., & Arensman, E. (2014). Risk factors for repetition of self-harm:

A systematic review of prospective hospital based studies. *PLoS One, 9*(1). Retrieved from http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0084282

Larson, J. L., Vos, C. M., & Fernandez, D. (2013). Interventions to increase physical activity in people with COPD: Systematic review. *Annual Review of Nursing Research, 31,* 297-326. doi:10.1891/0739-6686.31.297

Lejuez, C. W., Daughters, S. B., Nowak, J. A., Lynch, T., Rosenthal, M. Z., & Kosson, D. (2003). Examining the Inventory of Interpersonal Problems as a tool for conducting analogue studies of mechanisms underlying borderline personality disorder. *Journal of Behavior Therapy and Experimental Psychiatry, 34,* 313–324. doi:10.1016/j.jbtep.2003.11.002

Lewis, S. P., Lumley, M. N., & Grunberg, P. H. (2015). Early maladaptive schemas and non-suicidal self-injury among young adults: A preliminary investigation. *Counselling Psychology Quarterly, 28,* 386-403. doi:10.1080/09515070.2015.1074887

Liberati, A., Altman, D. G., Tetzlaff, J., Mulrow, C., Gøtzsche, P. C., Ioannidis, J. P. A., … Moher, D. (2009). The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: Explanation and elaboration. *Annals of Internal Medicine, 151,* 65-94. doi:10.7326/0003-4819-151-4-200908180-00136

Linehan, M. M. (1993). *Cognitive-behavioral treatment of borderline personality disorder.* New York, NY: Guilford.

Linehan, M. M., Armstrong, H. E., Suarez, A., Allmon, D., & Heard, H. (1991). Cognitive-behavioral treatment of chronically parasuicidal borderline patients. *Archives of General Psychiatry, 48,* 1060-1064. doi:10.1001/archpsyc.1991.01810360024003

Linehan, M. M., Comtois, K. A., Murray, M. A., Brown, M. Z., Gallop, R. J., Heard, H. L., … Lindenboim, N. (2006). Two-year randomized controlled trial and follow-up of dialectical behavior therapy vs therapy by experts for suicidal behaviors and borderline personality disorder. *Archives of General Psychiatry, 63,* 757-766. doi:10.1001/archpsyc.63.7.757

Linehan, M. M., Heard, H. L., & Armstrong, H. E. (1993). Naturalistic follow-up of a behavioral treatment for chronically parasuicidal borderline patients. *Archives of General Psychiatry, 50,* 971-974. doi:10.1001/archpsyc.1993.01820240055007

Linehan, M. M., Tutek, D. A., Heard, H. L., & Armstrong, H. E. (1994). Interpersonal outcome of cognitive behavioral treatment for chronically suicidal borderline patients. *The American Journal of Psychiatry, 151,* 1771-1776. doi:10.1176/ajp.151.12.1771

Lovibond, S. H., & Lovibond, P. F. (1995). *Manual for the Depression Anxiety Stress Scales* (2nd ed.). Sydney, Australia: Psychology Foundation of Australia.

Madge, N., Hawton, K., McMahon, E. M., Corcoran, P., De Leo, D., de Wilde, E. J., … Arensman, E. (2011). Psychological characteristics, stressful life events and deliberate self-harm: Findings from the Child & Adolescent Self-Harm in Europe (CASE) Study. *European Child and Adolescent Psychiatry, 20,* 499-508. doi:10.1007/s00787-011-0210-4

Martin, S., & Hammer, A. A. (1988). *Coping Resources Inventory*. Menlo Park, CA: Mind Garden.

McMain, S. F., Guimond, T., Barnhart, R., Habinski, L., & Streiner, D. L. (2016). A randomized trial of brief dialectical behaviour therapy skills training in suicidal patients suffering from borderline disorder. *Acta Psychiatrica Scandinavica, 135,* 138-148. doi:10.1111/acps.12664

National Institute for Health and Care Excellence. (2004). *Self-harm in over 8s: Short-*

term management and prevention of recurrence. Manchester, UK: Author. Retrieved from https://www.nice.org.uk/guidance/cg16

National Institute for Health and Care Excellence. (2011). *Self-harm in over 8s: Long-term management.* Manchester, UK: Author. Retrieved from https://www.nice.org.uk/guidance/cg133

Neacsiu, A. D., Rizvi, S. L., & Linehan, M. M. (2010). Dialectical behavior therapy skills use as a mediator and outcome of treatment for borderline personality disorder. *Behaviour Research and Therapy, 48,* 832-839. doi:10.1016/j.brat.2010.05.017

Nock, M. K. (2009). *Non-suicidal self-injury: Origins, assessment, and treatment*. Washington D. C., WA: American Psychological Association.

Ougrin, D., Tranah, T., Stahl, D., Moran, P., & Asarnow, J. R. (2015). Therapeutic interventions for suicide attempts and self-harm in adolescents: Systematic review and meta-analysis. *Journal of the American Academy of Child & Adolescent Psychiatry, 54,* 97-107. doi:10.1016/j.jaac.2014.10.009

Panos, P. T., Jackson, J. W., Hasan, O., & Panos, H. A. (2014). Meta-analysis and systematic review assessing the efficacy of dialectical behavior therapy (DBT). *Research on Social Work Practice, 24,* 213-223. doi:10.1177/1049731513503047

Patton, J. H., Stanford, M. S., & Barratt, E. S. (1995). Factor structure of the Barratt Impulsiveness Scale. *Journal of Clinical Psychology, 51,* 768–774. doi:10.1002/1097-4679(199511)51:6<768::aid-jclp2270510607>3.0.co;2-1

Pfohl, B., & Blum, N. (1997). *Borderline Evaluation of Severity over Time*. Iowa City, IA: University of Iowa.

Pfohl, B., Blum, N., St. John, D., McCormick, B., Allen, J., & Black, D. W. (2009). Reliability and validity of the Borderline Evaluation of Severity over Time

(BEST): A self-rated scale to measure severity and change in persons with borderline personality disorder. *Journal of Personality Disorders, 23,* 281–293. doi:10.1521/pedi.2009.23.3.281

Pollock, L. R., & Williams, J. M. (2004). Problem-solving in suicide attempters. *Psychological Medicine, 34,* 163-167. doi:10.1017/s0033291703008092

Rabin, R., & de Charro, F. (2001). EQ-5D: A measure of health status from the EuroQol group. *Annals of Medicine, 33,* 337–43. doi:10.3109/07853890109002087

Raj, M A. J., Kumaraiah, V., & Bhide, A. V. (2001). Cognitive-behavioural intervention in deliberate self-harm. *Acta Psychiatrica Scandinavica, 104,* 340-345. doi:10.1034/j.1600-0447.2001.00075.x

Rudd, M. D. (2006). Fluid vulnerability theory: A cognitive approach to understanding the process of acute and chronic suicide risk. In T. E. Ellis (Ed.), *Cognition and suicide: Theory, research, and therapy* (pp. 355-368). Washington D. C., WA: American Psychological Association.

Samoocha, D., Bruinvels, D. J., Elbers, N. A., Anema, J. R., & van der Beek, A. J. (2010). Effectiveness of web-based interventions on patient empowerment: A systematic review and meta-analysis. *Journal of Medical Internet Research, 12*(2). Retrieved from https://www.jmir.org/2010/2/e23/

Sheehan, D. V. (1983). *The anxiety disease.* New York, NY: Scribner.

Simons, J. S., & Gaher, R. M. (2005). The Distress Tolerance Scale: Development and validation of a self-report measure. *Motivation and Emotion, 29,* 83–102. doi:10.1007/s11031-005-7955-3

Smyth, J. M., & Arigo, D. (2009). Recent evidence supports emotion-regulation interventions for improving health in at-risk and clinical populations. *Current Opinion in Psychiatry, 22,* 205-210. doi:10.1097/YCO.0b013e3283252d6d

Spielberger, C. D. (1988) *Manual for the State-Trait Anger Expression Inventory*.

Odessa, FL: Psychological Assessment Resources.

Spielberger, C. D., Jacobs, G. A., Russell, S., & Crane, R. S. (1983). Assessment of anger: The State-Trait Anger Scale. In J. N. Butcher & C. D. Spielberger (Eds.), *Advances in personality assessment* (pp. 159-187). Hillsdale, NJ: Erlbaum.

Stewart, C., Quinn, A., Plever, S., & Emmerson, B. (2009). Comparing cognitive behaviour therapy, problem solving therapy, and treatment as usual in a high risk population. *Suicide and Life Threatening Behavior, 39,* 538-547. doi:10.1521/suli.2009.39.5.538

Townsend, E., Hawton, K., Altman, D. G., Erensman, E., Gunnell, D., Hazell, P., … van Heeringen, K. (2001). The efficacy of problem-solving treatments after deliberate self-harm: Meta-analysis of randomized controlled trials with respect to depression, hopelessness and improvement in problems. *Psychological Medicine, 31,* 979-988. doi:10.1017/s003329170105423

Turner, B. J., Austin, S. B., & Chapman, A. L. (2014). Treating nonsuicidal self-injury: A systematic review of psychological and pharmacological interventions. *Canadian Journal of Psychiatry, 59,* 576-585. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4244876/

Tyrer, P. (1990). Personality disorder and social functioning. In D. F. Peck & C. M. Shapiro (Eds.), *Measuring human problems: A practical guide* (pp. 119-142). Chichester, UK: Wiley.

Tyrer, P., Thompson, S., Schmidt, U., Jones, V, Knapp, M., Davidson, K., … Wessely, S. (2003). Randomized controlled trial of brief cognitive behaviour therapy versus treatment as usual in recurrent deliberate self-harm: The POPMACT study. *Psychological Medicine, 33,* 969-976. doi:10.1017/S0033291703008171

van Spijker, B. A. J., van Straten, A., & Kerkhof, A. J. F. M. (2014). Effectiveness of online self-help for suicidal thoughts: Results of a randomised controlled trial.

PloS One, 9(2). Retrieved from http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0090118

Waller, N. G., & Ross, C. A. (1997). The prevalence and biometric structure of pathological dissociation in the general population: Taxometric and behaviour genetic findings. *Journal of Abnormal Psychology, 106,* 499-510. doi:10.1037//0021-843x.106.4.499

Ware, J. E., & Sherbourne, C. D. (1992). The MOS 36-item Short-Form Health Survey (SF-36) I. Conceptual framework and item selection. *Medical Care, 30,* 473-483. doi:10.1097/00005650-199206000-00002

Washburn, J. J., Richardt, S. L., Styer, D. M., Gebhardt, M., Juzwin, K. R., Yourek, A., & Aldridge, D. (2012). Psychotherapeutic approaches to non-suicidal self-injury in adolescents. *Child and Adolescent Psychiatry and Mental Health, 6,* 14-21. doi:10.1186/1753-2000-6-14

Wei, S., Liu, L., Bi, B., Haiyan, L., Jinglin, H., Shanyong, T., … Ying, L. (2013). An intervention and follow-up study following a suicide attempt in the emergency departments of four general hospitals in Shenyang, China. *Crisis: The Journal of Crisis Intervention and Suicide Prevention, 34,* 107-115. doi:10.1027/0227-5910/a000181

Weissman, A. N., & Beck, A. T. (1992). Dysfunctional Attitude Scale. In. J. M. G. Williams (Ed.), *The psychological treatment of depression. A guide to the theory and practice of cognitive behaviour therapy* (2nd ed., pp. 100-104). New York, NY: Routledge, Chapman & Hall.

Weissman, M. M., & Bothwell, S. (1976). Assessment of social adjustment by patient self-report. *Archives of General Psychiatry, 33,* 1111-1115. doi:10.1001/archpsyc.1976.01770090101010

Williams, J. M. G., & Pollock, L. (2001). Psychological aspects of the suicidal process.

In K. van Heeringen (Ed.), *Understanding suicidal behaviour: The suicidal process approach to research, treatment and prevention* (pp. 76-94). Chichester, UK: Wiley.

World Health Organization. (2014). *Preventing suicide: A global imperative.* Geneva, Switzerland: Author. Retrieved from http://apps.who.int/iris/bitstream/10665/131056/1/9789241564779_eng.pdf?ua=1&ua=1

Zahl, D., & Hawton, K. (2004). Repetition of deliberate self-harm and subsequent suicide risk: Long-term follow-up study in 11583 patients. *British Journal of Psychiatry, 185,* 70-75. doi:10.1192/bjp.185.1.70

Zetterqvist, M. (2015). The DSM-5 diagnosis of nonsuicidal self-injury disorder: A review of the empirical literature. *Child & Adolescent Psychiatry & Mental Health, 9,* 31-43. doi:10.1186/s13034-015-0062-7

Zigmond, A. S., & Snaith, R. P. (1983). The Hospital Anxiety and Depression Scale. *Acta Psychiatrica Scandinavica, 67,* 361-370. doi:10.1111/j.1600-0447.1983.tb09716.x

Empirical report

Therapist effects and patient-level predictors of reliable and clinically significant change among at-risk patients: A multilevel modelling analysis

Abstract

Objectives: This study primarily aimed to evaluate the impact of patient-related variables and therapist effects on the likelihood of reliable and clinically significant improvement (RCSI) in psychological therapy among patients "at-risk" of self-injury.

Design: A quantitative cross-sectional design using a national practice-based dataset was employed.

Methods: Data were analysed for 4,976 patients treated across 81 therapists. All patients completed the Clinical Outcomes in Routine Evaluation-Outcome Measure (CORE-OM) pre- and post-therapy, and had positive CORE-OM risk-to-self scores. Logistic multilevel models of predictors of RCSI were developed.

Results: Approximately half (54.64%) of patients met criteria for RCSI. Lower pre-therapy CORE-OM non-risk scores, higher motivation, medication reductions, and being in employment/education predicted significantly higher odds of RCSI. Accounting for these variables, 10-13% of the variance in outcomes was attributable to therapist effects. Some therapists were significantly more, or less, effective than average.

Conclusions: Psychological therapy can effectively reduce distress among at-risk patients. Both patient-related variables and therapist effects significantly contribute to the likelihood of RCSI among at-risk patients, with therapist effects higher than those reported in studies of patients not at-risk. These results highlight the importance of measuring therapist effects in therapy outcome research, to ascertain accurate estimates of therapy effectiveness.

Practitioner Points

Clinical Implications:

- Psychological therapy can effectively reduce distress among at-risk patients.
- Recovery in therapy among at-risk patients may be facilitated by early

intervention and efforts to enhance patients' motivation and access to employment and education.

- Careful consideration is needed around allocating at-risk patients to therapists.
- Routine outcome measurements of therapist caseloads could inform supervisory processes, evaluation, and service development.

Limitations:

- The results may be less representative of patients with moderate-severe cognitive difficulties, patients accessing non-individual therapies, and patients who do not complete therapy.
- A single self-report measure assessed outcomes and risk.

The National Institute for Health and Care Excellence (NICE; 2011) recommends the provision of mental health services using a stepped care model, where treatments offered are the least restrictive for patients whilst being likely to be effective. The model is self-correcting, with outcomes being routinely monitored and patients referred to more or less intensive services as appropriate (Bower & Gilbody, 2005). Services at higher tiers of the model are proposed to be more adequately resourced to treat patients with more complex difficulties. Indeed, research suggests that adherence to stepped care models leads to better outcomes for patients (Gyani, Shafran, Layard, & Clark, 2013) which occur more rapidly (Oosterbaan et al., 2013) and cost-efficiently (Drummond et al., 2009).

A defining factor of complex difficulties might be risk. Patients "at-risk" of harming themselves are more likely to have comorbidities, substance misuse difficulties, and histories of trauma than patients not at-risk (Mann, Waternaux, Haas, & Malone, 1999) and benefit from more intense interventions (NICE, 2011). Accordingly, Barkham, Gilbert, Connell, Marshall, and Twigg (2005) found that, whilst pre-treatment symptom severity was comparable across 7,651 primary and secondary care National Health Service (NHS) patients, patients in secondary care reported significantly greater risk. Therefore, if the stepped care model is valid, at-risk patients should experience better treatment outcomes in higher tiers of the model.

However, many patients access treatment outside the NHS, including that offered by voluntary, private, and university/workplace services. There is comparatively little research with these populations and it is unclear how these services fit within a stepped care model. One study found similar rates of pre-treatment symptom severity and post-treatment recovery between NHS primary care patients and patients accessing voluntary and university/workplace services (Stiles, Barkham, & Wheeler, 2015). This suggests that voluntary and university/workplace services overlap with the low-intensity

services offered within NHS primary care, and may be less equipped to treat patients with more complex needs. However, few studies have assessed the prevalence of at-risk patients in these services and the outcomes that these populations experience.

Evaluating Outcomes

There is a paucity of therapy outcome research with at-risk patients. This population is frequently excluded from studies due to ethical and legal considerations (Linehan, 1997). Furthermore, the ability to draw firm conclusions from existing studies has been hampered by their heterogeneous methods, small samples, and inconsistent findings. Whilst several reviews have suggested promising treatment effects for psychological therapies in reducing patients' self-injury (Comtois & Linehan, 2006; Daigle, Pouliot, Chagnon, Greenfield, & Mishara, 2011), others have suggested null effects (Crawford, Thomas, Khan, & Kulinskaya, 2007; Hawton et al., 1999). In a recent meta-analysis of 45 randomised controlled trials (RCTs), Hetrick, Robinson, Spittal, and Carter (2016) found a significant treatment effect for psychosocial interventions in reducing self-injury and improving mental health outcomes among at-risk patients. However, the treatment effect was nonsignificant when only high quality studies were analysed. A criticism of this meta-analysis is that therapy was conflated with less clearly defined psychosocial interventions, thus the specific effect of therapy was unclear. Furthermore, therapies designed specifically for populations who are likely to self-injure (e.g., dialectical behaviour therapy) were not included. In a review of RCTs which did investigate the specific impact of therapies, including dialectical behaviour therapy, Hawton et al. (2016) concluded that therapy was associated with improvements in mental health symptoms, problem-solving, and reduced self-injury among at-risk patients.

Therefore, studies suggest that therapy can facilitate improvements in psychological wellbeing and self-injury outcomes among at-risk patients. However,

findings are inconsistent and studies have used heterogeneous methods with small and predominantly female samples. A further criticism of existing literature relates to the dominance of evidence-based practice (EBP) approaches, where RCT designs are adopted as the "gold standard" (Barkham & Mellor-Clark, 2003). The results of such studies are arguably less representative of routine practice due to the use of strict control criteria (McMillan & Morley, 2010). Alternatives to EBP approaches include practice-based evidence approaches, which evaluate outcomes in routine practice (see Barkham, Hardy, & Mellor-Clark, 2010). These two approaches form a reciprocal relationship, with the potential to develop practice recommendations grounded in both scientific empiricism and ecological validity.

Existing studies are further limited by their analyses of change. Traditionally, researchers have examined the statistical significance of changes in scores on standardised measures within or between therapies, and computed effect sizes to estimate change magnitude. Jacobson, Follette, and Revenstorf (1984) proposed that these analyses do not capture the reliability and clinical significance of change. Reliable change indices represent estimations of change independent of potential measurement error. Clinical significance is calculated from cut-off scores on measures which indicate whether patients are likely to belong to "clinical" or "nonclinical" populations. By applying these criteria, five categories of therapy outcome can be devised: (a) reliable and clinically significant improvement (RCSI), (b) reliable improvement only, (c) reliable and clinically significant deterioration, (d) reliable deterioration, and (e) no reliable change (Jacobson & Truax, 1991). Investigations of reliable and clinically significant changes in therapy among at-risk patients are rare. Three studies which adopted these criteria showed promising effects, with RCSI rates between 10.5% and 83.0% on psychological wellbeing measures among at-risk patients who completed therapy (Gratz & Gunderson, 2006; Gratz & Tull, 2011; Gratz, Tull, & Levy, 2013).

However, these studies investigated the effect of emotion regulation group therapy with women with personality disorder-related difficulties. Consequently, the proportion of at-risk patients who experience RCSI in different therapies, and for those with different genders and clinical presentations, remains unclear.

In summary, research suggests that therapy may facilitate improvements in self-injury and psychological wellbeing outcomes among at-risk patients. However, few studies have assessed the reliability and clinical significance of these outcomes in routine practice. Furthermore, as not all studies have found significant treatment effects, further investigation of the factors implicated in therapeutic change is necessary.

Predicting Outcomes

Research with diverse clinical populations has linked various patient-related variables to therapy outcomes. These include patient age, gender, employment, marital status, personality, symptom severity, comorbidity, and therapeutic alliance (Knopp, Knowles, Bee, Lovell, & Bower, 2013; Orlinsky, Rønnestad, & Willutzki, 2003; Wolitzky-Taylor, Arch, Rosenfield, & Craske, 2012). Comparatively fewer studies have investigated predictors of therapy outcomes with at-risk patients. Existing studies indicate that, in this population, treatment responses are associated with patient education, income, frequency of self-injury, comorbidity, and commitment and capacity to undertake therapy (Bedics, Atkins, Harned, & Linehan, 2015; Gratz, Dixon-Gordon, & Tull, 2014). However, the methods and outcomes used in studies are heterogeneous, with variables relating differently to outcomes based on different definitions of treatment response. Existing findings are also inconsistent. For example, greater pre-therapy psychopathology has been associated with both better (Gratz et al., 2014) and poorer (Guthrie et al., 2003) outcomes among at-risk patients. These factors complicate the ability to ascertain accurate predictors of outcome.

A further criticism of the existing literature is that few studies have assessed

how therapists themselves influence outcomes. The term *therapist effects* has been adopted to describe variability in outcomes between therapists, independent of the content of the therapy they provide and factors related to their patients. Beutler et al. (2003) suggested that therapist effects have been largely ignored due to the popularity of RCTs, which use control criteria to eliminate between-therapist variation when evaluating outcomes. However, several studies with at-risk patients have identified potential therapist effects. Husain et al. (2014) evaluated changes in suicidal ideation, depression, hopelessness, quality of life, and coping among 221 patients seen by 3 therapists. The results indicated no treatment effects among one therapist's patients, significant treatment effects on suicidal ideation and depression among the second therapist's patients, and significant treatment effects on all measures among the third therapist's patients. Thus, patient outcomes varied according to therapist allocation, despite all therapists being trained in fidelity to the therapeutic model. This study included predominantly female participants who accessed problem-solving therapy in Pakistan. Nonetheless, studies across other countries and therapeutic modalities have reported significant associations between therapist competence ratings and psychological wellbeing outcomes among at-risk patients (Davidson et al., 2004; Norrie, Davidson, Tata, & Gumley, 2013). These findings suggest a potential therapist effect on outcomes for at-risk patients.

The ability to disentangle therapist effects from the effects of patient-related variables on outcomes has been limited by analytical methods employed by studies. Researchers have traditionally evaluated therapy outcomes using analytical methods which assume that cases are independent (e.g., regression analyses). However, patients treated by one therapist may share similarities that patients treated by another therapist do not share. Therefore, patients are not independent from other patients seen by the same therapist. This creates a hierarchical structure, where patients are nested within

therapists. To accurately assess therapist effects, analytical methods which model hierarchical data must be employed. One such method is multilevel modelling. Multilevel modelling explicitly models dependence within data and can therefore estimate the relative influence of therapist effects and patient-related variables on outcomes. Previous studies which have adopted this method have found that, when accounting for other variables, therapist effects account for between 1% and 10% of the variance in therapy outcomes (Firth, Barkham, Kellett, & Saxon, 2015; Green, Barkham, Kellett, & Saxon, 2014; Saxon & Barkham, 2012). However, no studies have used multilevel modelling to estimate the relative impact of patient-related variables and therapist effects on outcomes among at-risk patients.

Current Study

Rationale. Identifying factors which predict RCSI among at-risk patients is important for the development of effective therapies. Such information could enable at-risk patients to be signposted to services most adequately resourced to meet their needs, or enable services to develop their responsiveness to these needs. Identification of patient-related predictors of RCSI could enable therapists to modify their practice in ways that enhance the potential for beneficial therapy outcomes. Furthermore, estimations of therapist effects in this area could have important implications for the training, supervision, and management of clinicians.

Aims and hypotheses. The current study had three main aims. First, it aimed to investigate variability in risk across NHS, voluntary, private, and university/workplace services. Specifically, the risk-to-self scores of NHS primary care, secondary care, and tertiary care patients were compared. Risk-to-self scores were expected to be lowest among primary care patients and highest among tertiary care patients, in line with the stepped care model. The risk-to-self scores of patients accessing NHS primary care, voluntary, and university/workplace services were also compared, to determine whether

there were significant differences in risk-to-self across these sectors. The second aim was to investigate rates of reliable and clinically significant change among at-risk patients in therapy. No specific hypotheses were made in relation to rates of reliable and clinically significant change due to the lack of existing research. Thirdly, this study aimed to identify predictors of RCSI for at-risk patients in therapy. In line with previous research, therapist effects were expected to account for up to 10% of the variance in outcomes. No hypotheses were made regarding the significance of specific patient-related predictors due to inconsistencies in previous findings.

Method

A quantitative cross-sectional design was employed using data from the Clinical Outcomes in Routine Evaluation Practice-Based Evidence National Database-2011. Ethical approval for this study was covered by the East Leeds Research Ethics Committee (REC reference: 05/Q1206/128; see Appendix B). Below, the original dataset, study-specific dataset, measures, study variables, and analyses are described.

Original Dataset

The original dataset included data for 104,747 patients treated across 2,442 therapists. Information was available from 15 voluntary, 18 university/workplace, 7 primary care, 8 secondary care, 2 tertiary care, and 2 private services. Patients were referred to services between October 1994 and December 2011.

Study-Specific Dataset

Data were included in the current study for patients aged ≥ 16 years who were accepted for therapy and had complete pre- and post-therapy outcome measures. Data were excluded for patients with moderate or severe cognitive impairments and for patients who received group, couples, or family therapy. This resulted in a study-specific dataset comprising data for 29,277 patients treated across 1,631 therapists. Data from the whole study-specific dataset were analysed to investigate variability in risk

across services. Each following analysis included subsets of the study-specific dataset, with the final multilevel model being developed using data from 4,976 patients (see the Analyses section, for more information).

Demographic and clinical information for excluded patients, the study-specific dataset, and the sample with which the multilevel model was developed are presented in Table 1 (for continuous variables) and Table 2 (for categorical variables). No information was available about therapists. Significant between-group differences were investigated using independent *t*-tests, with Cohen's *d* effect sizes, and Pearson's chi-square analyses, with Cramer's *V* effect sizes.

As Tables 1 and 2 show, patients in the study-specific dataset had a mean (*SD*) age of 37.84 (12.82) years and were predominantly female, White, and in employment or education. The most common presenting problems were anxiety, depression, and interpersonal difficulties. The majority of patients accessed university/workplace, primary care, and voluntary services. Regarding therapeutic provisions, patients most commonly received integrative therapies with planned endings.

Significant between-group differences emerged (*p*s < .001) between the study-specific dataset and excluded patients on all variables presented in Tables 1 and 2. Relative to the study-specific dataset, excluded patients were significantly younger (*d* = 0.20); had lower therapy attendance (*d* = 0.65); had less integrative therapies (*V* = .10); Table 1.

Descriptive Statistics for Excluded Patients, the Study-Specific Dataset, and the Sample with Which the Multilevel Model was Developed on Continuous Variables

Variable	Excluded cases			Study-specific dataset			MLM sample		
	n	*M*	*SD*	*n*	*M*	*SD*	*n*	*M*	*SD*
Age (years)	69,587	35.21	13.00	29,149	37.84	12.82	4975	38.46	13.07
Number of presenting problems	56,274	4.21	3.21	27,035	3.63	1.76	4976	3.94	1.80
Therapy attendance	35,314	0.76	0.26	27,470	0.90	0.16	4936	0.90	0.15

Note. MLM sample = sample with which the final multilevel model was developed.

Table 2.

Descriptive Statistics for Excluded Patients, the Study-Specific Dataset, and the Sample with Which the Multilevel Model was Developed on Categorical Variables

Variable	Category	Excluded cases		Study-specific dataset		MLM sample	
		n	Frequency (%)	*n*	Frequency (%)	*n*	Frequency (%)
Gender	Female	73,806	66.21	29,276	68.65	4,976	68.23
	Male		33.79		31.35		31.77
Employment	Employed/student	60,352	75.63	27,625	81.88	4,976	78.46
	Unemployed		21.35		15.00		17.89
	Retired		3.02		3.13		3.66
Problem type(s)	Depression	56,274	60.80	27,035	57.62	4,976	66.78
	Anxiety/stress		72.28		75.69		76.73
	Psychosis		7.18		0.60		0.50
	Personality problems		14.25		3.40		2.35
	Cognitive/learning		7.90		0.80		0.68
	Eating disorder		9.62		3.60		4.16
	Physical problems		18.26		14.70		17.77
	Addiction		12.43		4.88		5.99
	Trauma/abuse		24.88		17.57		20.92
	Loss/bereavement		30.06		30.10		34.95
	Self-esteem		44.48		46.71		51.25
	Interpersonal		56.06		53.92		55.69
	Welfare/living		19.71		14.59		17.73
	Work/academic		30.26		31.50		33.62
	Other		12.55		7.04		5.35
Ethnicity	Asian	64,670	6.23	29,276	4.00	4,976	5.08
	Mixed race		0.95		0.68		0.72
	Black		5.33		3.08		3.24
	White		83.42		83.80		87.52
	Other/not stated		4.07		8.43		3.44
Therapy modality	Integrative	34,526	50.23	27,485	60.05	4,971	69.08
	Other		49.77		39.95		30.92
Therapy ending	Planned	35,789	48.93	27,754	91.79	4,968	92.83
	Unplanned		51.07		8.21		7.17
Service type	Voluntary	75,199	20.20	29,277	22.60	4,976	5.25
	University/workplace		41.19		42.44		41.86
	Primary care		24.39		29.62		50.68
	Secondary care		8.23		3.96		0.97
	Tertiary care		5.79		0.39		0.00
	Private		0.20		0.98		1.25

Note. MLM sample = sample with which the final multilevel model was developed. Patients could report multiple problem types. Frequencies may not sum to 100% due to rounding.

had fewer planned therapy endings (V = .45); accessed less primary care and private services, and more secondary care and tertiary care services (V = .16); were more likely to have Asian, Black, or "other" ethnicities, and less likely to be mixed race (V = .11); and were more likely to report difficulties with psychosis, personality, cognition, eating, and addiction (V = .15). Effect sizes for differences on other variables were very small (d = .005 and Vs ≤ .08).

Among the study-specific dataset, significant between-group differences emerged (ps < .01) between patients who were included in the multilevel modelling analysis and patients who were not included regarding age, number of presenting problems, ethnicity, employment, problem types, therapy modality, service use, and therapy endings. Relative to patients not included in this analysis, patients who were included reported significantly more problems (d = 0.38), and accessed less voluntary, secondary care, and tertiary care services, and more primary care services (V = .25). Effect sizes for differences on other variables were very small (ds ≤ .06 and Vs ≤ .09). The full results of these analyses can be found in Appendix C.

Measures

Clinical Outcomes in Routine Evaluation-Outcome Measure (CORE-OM).

All patients had completed the CORE-OM (Barkham et al., 1998, 2001; Evans et al., 2002; see Appendix D) pre- and post-therapy. The CORE-OM is a standardised self-report measure of psychological distress. The measure features 34 statements related to psychological wellbeing which respondents rate their experience of on a 5-point Likert scale ranging from 0 (*not at all*) to 4 (*most or all of the time*). The maximum score is 40, with higher scores indicating more distress. Scores can also be computed for four subscales which relate to wellbeing (4 items), symptoms (12 items), functioning (12 items), and risk (6 items). Within the risk subscale, 4 items measure risk-to-self and 2 items measure risk-to-others. Pre-post changes in CORE-OM scores were used to

measure therapy outcomes, and the risk subscale was used to measure risk.

The validity of the CORE-OM is well documented (see Barkham, Mellor-Clark, Connell, & Cahill, 2006, for a review). Research has demonstrated the psychometric validity of using the risk and non-risk items as distinct measures of risk and psychological distress, respectively (Lyne, Barrett, Evans, & Barkham, 2006). All subscales have good internal reliability, with Cronbach alpha values from 0.70 to 0.97 across clinical populations (Barkham et al., 2005). Studies have also found that the measure has good test-retest reliability, good sensitivity to change, convergent validity with other practitioner- and patient-completed measures, and a good ability to discriminate between clinical and nonclinical populations (Barkham et al., 2001; Connell et al., 2007; Evans et al., 2002).

Clinical Outcomes in Routine Evaluation-Therapy Assessment Form (CORE-TAF). The CORE-TAF (Mellor-Clark & Barkham, 2000; Mellor-Clark, Barkham, Connell, & Evans, 1999; see Appendix E) was used to construct predictor variables. Therapists completed the CORE-TAF for each patient prior to commencing therapy. This information sheet records data regarding the patient's referral, demographic information, relationships, service use, medication use, presenting problems, risk, and coping strategies.

Clinical Outcomes in Routine Evaluation-End of Therapy Form (CORE-ETF). The CORE-ETF (Mellor-Clark & Barkham, 2000; Mellor-Clark et al., 1999; see Appendix F) was used to construct predictor variables. Therapists completed the CORE-ETF for each patient at the end of therapy. This information sheet records data regarding the patient's therapeutic provision, therapy engagement, presenting problems, medication changes, and the perceived benefits of therapy.

Study Variables

Outcome. A binary outcome variable was constructed using pre- and post-

therapy CORE-OM scores. Participants were categorised as having met criteria for RCSI or not, according to Jacobson and Truax's (1991) criteria. Changes of ≥ 5 points on the CORE-OM were considered reliable (Barkham et al., 2006). Clinical significance was determined using a cut-off score of 10, which has been found to discriminate between clinical and nonclinical populations with 88% specificity and 87% sensitivity (Connell et al., 2007).

Predictor variables. The CORE-OM, CORE-TAF, and CORE-ETF provided a substantial number of potential predictor variables. Variables were excluded where (a) they appeared irrelevant, (b) all patients scored the same, (c) most patients (≥ 85%) had missing data, (d) information regarding the variable was provided elsewhere, (e) they represented an outcome other than CORE-OM score, and (f) their validity was unclear due to inconsistently applied scoring criteria. The remaining variables were used as independent predictors or were used to create predictors.

Analyses

Analyses were conducted using IBM SPSS Statistics Version 23.0 (IBM Corporation, 2015) and MLwiN Version 2.36 (Rasbash, Charlton, Browne, Healy, & Cameron, 2016). Due to the large sample size, marginal effects and deviations from normality were expected to be statistically significant. Consequently, an alpha level of < .001 was adopted across all analyses and statistical tests of data normality were considered inappropriate (Field, 2009). Where quantile-quantile plots and histograms suggested non-normally distributed data, nonparametric analyses were conducted. In each instance, the results of the parametric and nonparametric analyses were largely equivalent. Indeed, research indicates that *t*-tests and analyses of variance (ANOVAs) are robust to violations of normality in large samples (Field, 2009; Lumley, Diehr, Emerson, & Chen, 2002). Therefore, the results of parametric analyses are presented in this report and the results of nonparametric analyses can be found in Appendix G.

Details of analyses for each study aim are described below.

Analysis of variability in risk and non-risk scores across services. Using the whole study-specific dataset, differences in CORE-OM risk-to-self and non-risk scores between patients in different service types were investigated using descriptive statistics and ANOVAs with eta-squared effect sizes. Games-Howell post hoc comparisons were computed to explore significant differences between services due to heterogeneity of variance in data.

Analysis of reliable and clinically significant change among at-risk patients. Data for patients with CORE-OM risk-to-self scores of 0 were excluded (n = 15,331). Descriptive statistics were computed to explore pre- and post-therapy CORE-OM scores for the remaining 13,946 at-risk patients. To estimate magnitude of change, Cohen's d effect sizes were computed using both the non-pooled standard deviation and Morris and DeShon's (2002) correction for dependence between the pre-post therapy mean scores. Patients were categorised as having met criteria for RCSI, reliable improvement, no reliable change, reliable deterioration, or reliable and clinically significant deterioration. The frequencies of these change categories were explored using descriptive statistics and Pearson's chi-square analysis.

Analysis of predictors of RCSI for at-risk patients. Patients with pre-therapy CORE-OM scores < 10 were excluded (n = 229), due to their inability to meet criteria for RCSI. The remaining 13,717 patients were categorised as having met criteria for RCSI or not. Each variable's potential to predict RCSI was analysed prior to developing a model of predictors of RCSI. Independent t-tests and Cohen's d effect sizes were computed to determine whether scores on continuous variables significantly differed between patients who did and did not meet criteria for RCSI. Pearson's chi-square analyses and Cramer's V effect sizes were computed to determine whether scores on categorical variables significantly differed between patients who did and did not meet

criteria for RCSI. Due to the large sample size, marginal effects were statistically significant. Consequently, variables were selected for entry into the model based on Cohen's (1988) conventional benchmarks for small effects (i.e., ds ≥ 0.2 and Vs ≥ .1). This enabled the selection of relevant predictors to streamline the modelling process and produce a parsimonious model. Where categorical variables had ≥ 3 categories, their categories were refined based on sample sizes and adjusted residuals (see Figure 1). Both original and refined categorical variables were assessed for entry to the model. Where refining categories did not impact the outcome, the parsimonious variables were retained. The results of analyses on original categorical variables can be found in Appendix H.

Variable	Original categories	Refined categories
Employment	Employed/student.	Employed/student
	Not employed/student. Retired.	Unemployed
Concurrent care	None.	No
	Voluntary. Primary. Secondary. Specialist.	Yes
Previous care	None.	No
	Voluntary. Primary. Secondary. Specialist.	Yes
Ethnicity	White.	White
	Asian. Mixed race. Black. Other/not stated.	Other
Therapy frequency	Weekly.	Weekly
	Less than weekly. More than weekly. No set frequency.	Other
Living status	Alone.	Alone
	With friends/family.	With friends/family
	Shared housing. Temporary housing. Institution.	Other
Caring role	None.	No
	Caring for children. Other full/part-time caring role.	Yes
Medication use during therapy	None.	None
	Reduced. Discontinued.	Reduced
	Modified. Increased. Started.	Increased/modified
	Maintained.	Maintained

Figure 1. Original and refined categorical variables.

Multilevel modelling. To identify predictors of whether at-risk patients met

criteria for RCSI, a two-level logistic multilevel model was developed using iterative generalised least squares estimation procedures (see Rasbash, Steele, Browne, & Goldstein, 2016). The impact of patient-related variables and therapist effects was assessed. Patients were entered at level one of the model and were grouped within therapists, who were entered at level two. Due to a failure to meet sample size requirements, a three-level model in which therapists were grouped within services was not possible to compute.

Sample size and power. Traditional power analyses may be unreliable for multilevel modelling (Field, 2009). For accurate investigations of therapist effects Maas and Hox (2005) recommend a minimum therapist sample size of 50, whereas Soldz (2006) recommends a minimum therapist sample size of 30 with each therapist providing data for ≥ 30 patients. Patients with missing data on predictor variables were excluded from analyses to meet requirements for multilevel modelling (n = 8,741). Consequently, data were available for the multilevel modelling analysis for 8,885 patients treated across 1,020 therapists. Eighty-one therapists provided data for ≥ 30 patients. Therefore, both Maas and Hox's (2005) and Soldz's (2006) sample size recommendations were satisfied. The multilevel model was developed using data from the 81 therapists who each provided data for ≥ 30 patients (patient n = 4,976). To test the robustness of the final multilevel model, the whole multilevel modelling analysis was re-run using data from the larger sample of 1,020 therapists (patient n = 8,885).

Multicollinearity analyses. Analyses were undertaken prior to multilevel modelling to identify multicollinearity between predictor variables. Tolerance and variance inflation factor values were calculated for each predictor. The strength of relationships between each pair of predictors was estimated using ANOVAs (with eta-squared effect sizes), correlation analyses, and Pearson's chi-square analyses (with Cramer's V effect sizes). Variables were considered for exclusion if they had

large-sized significant relationships with another variable.

Single-level model analyses. First, a single-level model was developed to identify patient-level predictors of the likelihood of patients meeting criteria for RCSI. Predictor variables which had been found to significantly differentiate between patients who did and did not meet criteria for RCSI with effect sizes of $d \geq 0.2$ or $V \geq .1$ were added to the model, one by one, in order of their expected contribution. Continuous predictors were grand-mean centred prior to entry to safeguard against multicollinearity (Hofman & Gavin, 1998). For categorical predictors with ≥ 3 categories, Wald tests were used to assess whether categories significantly differed in their contribution to the model. Where the contributions of categories did not significantly differ, categories were collapsed and the model was re-run with the refined variable.

Predictors with *z* scores ≥ 3.29 (where $z = \beta / SE$) were considered significant and nonsignificant predictors were removed. Odds ratios were calculated by taking exponentials of beta coefficients. For continuous predictors, the odds of meeting criteria for RCSI multiply by the odds ratio for each unit increase in the variable. For categorical predictors, the odds of meeting criteria for RCSI multiply by the odds ratio from one category to the next. Odds ratios > 1 indicated higher odds of meeting criteria for RCSI and odds ratios < 1 indicated lower odds of meeting criteria for RCSI.

Two-level model analyses. In the second step of the multilevel modelling analysis, therapists were added to the model and changes to the model were assessed. As therapists represent random effects (Crits-Cristoph, Tu, & Gallop, 2003), each therapist's regression line and intercept were allowed to vary. This enabled the likelihood of a patient meeting criteria for RCSI to vary between therapists. Predictors which were nonsignificant in the single-level model analysis were re-entered to assess any change to their predictive ability. Categorical predictors which had been refined in the single-level model analysis were re-entered in their original form. Wald tests were

again used to collapse categories of categorical predictors that did not significantly differ in their contributions to the model. The model was checked for random slopes. This enabled the effects of patient-level predictors to vary between therapists.

The model was initially developed using first order marginal quasi-likelihood (MQL) approximation procedures. As this type of approximation can lead to downwards-biased estimates, the final model was also run with second order penalised quasi-likelihood (PQL) approximation procedures. Both the MQL and PQL estimates are presented in the Results section. In the final model, analyses were undertaken to determine whether there were significant interaction effects between predictors and to determine whether continuous predictors had curvilinear relationships with the outcome. Standardised residuals were analysed to explore assumptions of normality and homoscedasticity.

Therapist effect estimates. The variance partition coefficient (VPC) was calculated to estimate therapist effects. The VPC represented the proportion of total residual variance in the model that was attributable to therapists. The VPC was estimated using the linear threshold model, where therapist-level variance (σ_{u0}^2) was divided by therapist-level variance plus estimated patient-level variance set at 3.29 (Snijders & Bosker, 1999). The Wald test was used to determine the significance of the VPC. This is an approximate measure of therapist effects, which indicates if there are significant unexplained differences between therapists in outcomes, as variance estimates are not normally distributed.

Public and Patient Involvement

The collection of study data was largely shaped by the involvement of therapists and patients, who volunteered data for the purposes of research to improve service delivery. Service user groups were consulted regarding the current report's terminology. Unfortunately, only one response was received. This response suggested that the most

acceptable term to describe patients considered at-risk of harming themselves was "at-risk". This term is adopted throughout this report.

Results

The results of each analysis are presented below. First, analyses of differences in risk and non-risk scores between services are presented. Second, rates of reliable and clinically significant change among at-risk patients are presented. Finally, analyses are presented for each step in the development of a model of predictors of RCSI in therapy among at-risk patients.

Variability in Risk and Non-Risk Scores Across Services

Table 3 presents descriptive statistics for pre-therapy CORE-OM risk-to-self and non-risk scores across service types. A one-way between-subjects ANOVA indicated a significant difference in risk-to-self scores across services with a small effect size, $F(5, 29271) = 135.70$, $p < .001$, $\eta^2 = .02$. Games-Howell post hoc comparisons suggested that patients in secondary care reported significantly higher risk-to-self than patients in any other service type. Patients in university/workplace services reported significantly lower risk-to-self than patients in voluntary and primary care services. No other significant between-group differences emerged.

Table 3.

Descriptive Statistics for CORE-OM Risk-to-Self and Non-Risk Scores Across Services

		Pre-therapy risk-to-self scores		Pre-therapy non-risk scores	
Service type	*n*	*M*	*SD*	*M*	*SD*
Voluntary	6,617	4.85	7.28	19.89	7.19
University/workplace	12,426	3.97	6.76	20.33	6.73
Primary care	8,672	4.90	7.43	21.42	6.82
Secondary care	1,159	9.65	10.48	23.04	7.53
Tertiary care	115	5.72	7.17	20.78	7.01
Private	288	3.79	6.57	20.46	7.09

A one-way between-subjects ANOVA suggested that CORE-OM non-risk scores also significantly differed between services, $F(5, 29271) = 71.28$, $p < .001$,

although the effect size was very small ($\eta^2 = .01$). Games-Howell post hoc comparisons suggested that patients in secondary care had significantly higher non-risk scores than patients in voluntary, university/workplace, primary care, and private services. Patients in voluntary services had significantly lower non-risk scores than patients in university/workplace and primary care services, and patients in university/workplace services had significantly lower non-risk scores than patients in primary care.

Rates of Reliable and Clinically Significant Change Among At-Risk Patients

Among patients with CORE-OM risk-to-self scores > 0 (n = 13,946), total CORE-OM scores reduced from pre-therapy (M = 21.26, SD = 5.34), to post-therapy (M = 10.64, SD = 7.06). Both the non-pooled effect size (d = 2.0) and the effect size calculated using Morris and DeShon's (2002) correction (d = 1.5) indicated large effects. Figure 2 displays the frequencies of reliable and clinically significant change categories among at-risk patients.

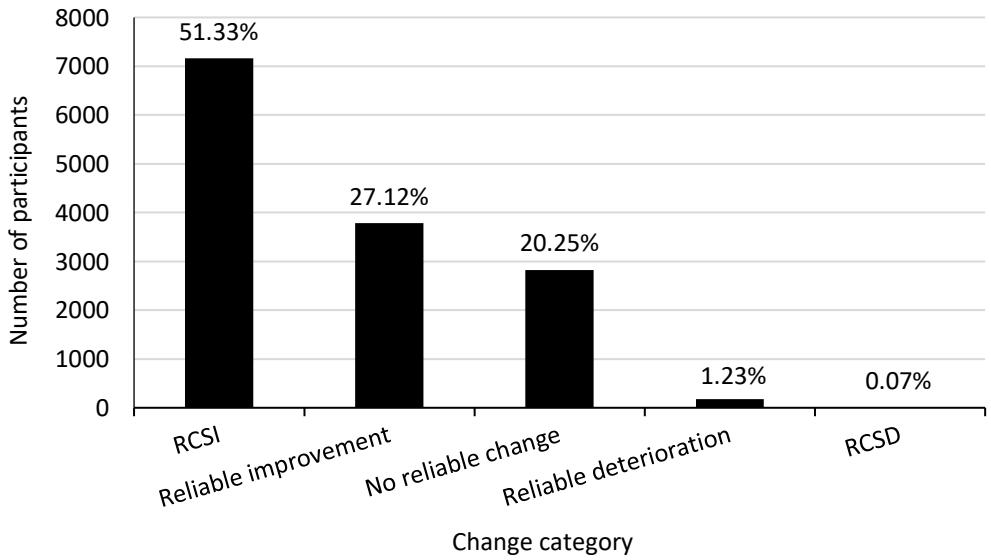

Figure 2. Bar chart depicting frequencies of reliable and clinically significant change categories in pre-post therapy Clinical Outcomes in Routine Evaluation-Outcome Measure scores among at-risk patients. RCSI = reliable and clinically significant improvement. RCSD = reliable and clinically significant deterioration.

Pearson's chi-square analysis indicated a significant difference between proportions of reliable and clinically significant change categories, $\chi^2(4) = 12421.83$,

$p < .001$. Frequencies indicated that, relative to a normal distribution, more patients met criteria for RCSI and reliable improvement and less patients met criteria for clinically significant and/or reliable deterioration.

Predictors of RCSI for At-Risk Patients

Across at-risk patients with pre-therapy CORE-OM scores above the clinical cut-off, 7,158 patients (52.18%) met criteria for RCSI and 6,559 patients (47.82%) did not (i.e., "No RCSI"). Differences between the scores of patients who did and did not meet criteria for RCSI on potential predictor variables were analysed using independent *t*-tests (see Table 4) and Pearson's chi-square analyses (see Table 5). Only results for variables with effect sizes of $d \geq 0.2$ and $V \geq .1$ are presented due to space constraints (see Appendix I for the full results).

Table 4 shows that patients who met criteria for RCSI had significantly lower pre-therapy CORE-OM risk and non-risk scores than patients who did not meet criteria for RCSI. Effect sizes for these differences were small-medium.

Table 4.

Independent t-tests Comparing Data on Continuous Predictors Between Patients Who Did and Did Not Meet Criteria for RCSI

	No RCSI		RCSI				
Predictor	*M*	*SD*	*M*	*SD*	*t*	df	*d*
Pre-therapy CORE-OM risk score[a]	8.59	6.62	6.71	5.50	17.95*	12788.71	0.31
Pre-therapy CORE-OM non-risk score	25.59	5.42	23.42	5.33	23.65*	13715.00	0.40

Note. RCSI = reliable and clinically significant improvement. *d* = Cohen's *d*. CORE-OM = Clinical Outcomes in Routine Evaluation-Outcome Measure.

[a] equal variances not assumed.

* $p < .001$.

Table 5 indicates that significant between-group differences emerged, with small effect sizes, regarding medication, problem duration, motivation, psychological mindedness, working alliance, and employment status. Frequencies and adjusted standardised residuals indicated that there were significant positive associations

between patients meeting criteria for RCSI and having medication reductions; good motivation, psychological mindedness, and working alliances; shorter problem durations; and being in employment or education.

Table 5.

Pearson's Chi-Square Analyses Comparing Data on Categorical Predictors Between Patients Who Did and Did Not Meet Criteria for RCSI

		No RCSI		RCSI				
Variable	Category	Obs.	Exp.	Obs.	Exp.	χ^2	df	*V*
Medication use in therapy	None	2,724	2,855.9	3,358	3,226.1	220.64[a]	3	.14[a]
	Reduced	249	413.7	632	467.3			
	Increased/modified	443	357.8	319	404.2			
	Maintained	1,993	1,781.6	1,801	2,012.4			
Employment	Employed/student	4,448	4,764.0	5,571	5,255.0	178.13[b]	1	.12[a]
	Unemployed	1,694	1,378.0	1,204	1,520.0			
Motivation	Poor	402	248.6	124	277.4	591.97[a]	2	.23[a]
	Moderate	1,742	1,327.4	1,067	1,481.6			
	Good	3,098	3,666.0	4,660	4,092.0			
Working alliance	Poor	247	150.2	71	167.8	498.84[a]	2	.21[a]
	Moderate	1,657	1,247.8	984	1,393.2			
	Good	3,333	3,838.9	4,792	4,286.1			
Psychological mindedness	Poor	655	455.9	311	510.1	477.65[a]	2	.21[a]
	Moderate	2,199	1,854.6	1,731	2,075.4			
	Good	2,372	2,915.5	3,806	3,262.5			
Problem duration	< 6 months	570	712.8	942	799.2	114.37[a]	3	.10[a]
	6-12 months	483	568.5	723	637.5			
	> 12 months	1,262	1,269.1	1,430	1,422.9			
	Recurring/continuous	3,164	2,928.5	3,048	3,283.5			

Note. RCSI = reliable and clinically significant improvement. Obs. = Observed. Exp. = Expected. *V* = Cramer's *V*.
[a] Monte Carlo $p < .001$ with 99% confidence intervals [.00, .00]. [b] Exact $p < .001$.

Multilevel modelling. The multilevel modelling analysis included data for 4,976 patients treated across 81 therapists. Among this sample, 2,719 patients met criteria for RCSI (54.64%) and 2,257 did not (45.36%).

Multicollinearity analyses. Pearson's chi-square analyses indicated large-sized significant relationships between ratings of psychological mindedness and working alliance, $\chi^2(2) = 4859.63$, $p < .001$, $V = .50$; psychological mindedness and motivation, $\chi^2(2) = 6986.37$, $p < .001$, $V = .56$; and working alliance and motivation, $\chi^2(2) =$

4870.55, $p < .001$, $V = .47$. Among these variables, motivation had the largest effect size in discriminating patients who did and did not meet criteria for RCSI (see Table 5). Consequently, motivation was retained as a potential predictor, and psychological mindedness and working alliance were not entered in the model. No other multicollinearity issues were identified (see Appendix J).

Single-level model. Table 6 presents the single-level model developed with patient-level predictors of meeting criteria for RCSI (see Appendix K for the full model). The results of Wald test analyses led to the collapsing of categories of *poor* and *moderate* motivation, $\chi^2(1) = 9.43$, $p = .002$, and *increased/modified* and *maintained* medication, $\chi^2(1) = 2.53$, $p = .112$. The contribution of the *> 12 months* problem duration category did not significantly differ from the contributions of either the *6-12 months* problem duration category, $\chi^2(1) = 4.62$, $p = .03$, or the *recurrent/continuous* problem duration category, $\chi^2(1) = 4.85$, $p = .03$. From a theoretical perspective, experiencing a mental health difficulty for twelve months may be very different from experiencing a life-long mental health difficulty. Consequently, these categories were refined to represent problem durations of *> 6 months* and *recurrent/continuous*.

Table 6.

Single-Level Model of Patient-Level Predictors of Meeting Criteria for RCSI

Variable	β	*SE*	*z*	*OR*	95% CI for *OR*
Pre-therapy CORE-OM non-risk score	-0.08	0.01	-12.67	0.92	[0.91, 0.94]
Reduced medication[a]	0.93	0.12	7.58	2.53	[2.00, 3.21]
Increased/maintained medication[a]	-0.10	0.06	-1.48	0.90	[0.80, 1.02]
Poor/moderate motivation[b]	-0.96	0.07	-13.71	0.38	[0.33, 0.44]
Problem duration >6 months[c]	-0.29	0.10	-3.02	0.75	[0.62, 0.91]
Problem duration – continuous/recurrent[c]	-0.52	0.09	-5.66	0.59	[0.50, 0.71]

Note. OR = odds ratio. CI = confidence interval. CORE-OM = Clinical Outcomes in Routine Evaluation-Outcome Measure.

[a] reference category *no medication.* [b] reference category *good* motivation. [c] reference category *< 6 months* problem duration.

Table 6 shows that the odds of meeting criteria for RCSI were significantly lower for patients with higher pre-therapy CORE-OM non-risk scores, poor/moderate

motivation as opposed to good motivation, and recurrent/continuous difficulties as opposed to difficulties with shorter durations. The odds of meeting criteria for RCSI were significantly higher for patients with medication reductions during therapy. Nonsignificant predictors that were removed from the model included pre-therapy CORE-OM risk scores (z = -1.80) and employment status (z = -2.75).

Two-level model. Table 7 shows the final two-level model developed with patient-level predictor variables at level one and therapists at level two (see Appendix L for the full model). The results of Wald test analyses again led to the collapsing of categories of *poor* and *moderate* motivation, $\chi^2(1) = 9.34$, $p = .002$, and *increased/modified* and *maintained* medication, $\chi^2(1) = 2.02$, $p = .156$.

Table 7.

Two-Level Model of Patient-Level Predictors and Therapist Effects in Predicting the Likelihood of Meeting Criteria for RCSI

Variable	β	*SE*	*z*	*OR*	95% CIs for OR
MQL estimates					
Pre-therapy CORE-OM non-risk score	-0.07	0.01	-12.00	0.93	[0.91, 0.95]
Reduced medication[a]	0.75	0.13	5.95	2.12	[1.64, 2.73]
Increased/maintained medication[a]	-0.11	0.07	-1.67	0.90	[0.78, 1.03]
Poor/moderate motivation[b]	-1.04	0.08	-13.81	0.35	[0.30, 0.41]
Unemployed[c]	-0.42	0.08	-4.99	0.66	[0.56, 0.77]
PQL estimates					
Pre-therapy CORE-OM non-risk score	-0.08	0.01	-13.17	0.92	[0.91, 0.94]
Reduced medication[a]	0.84	0.13	6.28	2.32	[1.80, 2.99]
Increased/maintained medication[a]	-0.13	0.07	-1.81	0.88	[0.77, 1.01]
Poor/moderate motivation[b]	-1.15	0.08	-14.56	0.32	[0.27, 0.37]
Unemployed[c]	-0.47	0.09	-5.32	0.63	[0.52, 0.75]

Note. MQL = marginal quasi-likelihood. *OR* = odds ratio. CI = confidence interval. CORE-OM = Clinical Outcomes in Routine Evaluation-Outcome Measure. PQL = penalised quasi-likelihood.

[a] reference category *no medication*. [b] reference category *good* motivation. [c] reference category *employed/student*.

Table 7 shows that the odds of meeting criteria for RCSI were significantly lower for patients with higher pre-therapy CORE-OM non-risk scores, poor/moderate motivation as opposed to good motivation, and for patients who were unemployed as opposed to being employed or in education. The odds of meeting criteria for RCSI were

significantly higher for patients with medication reductions during therapy. Concordant with the single-level model, pre-therapy CORE-OM risk scores did not significantly contribute to the two-level model (z = -0.83). In contrast to the single-level model, employment status became a significant predictor and problem duration was no longer a significant predictor in either its refined or original form (zs ≤ -2.94).

No random slopes were identified (ps ≥ .19), thus the effects of patient-level predictors did not significantly vary between therapists. No significant interaction effects were found between predictors (zs ≤ 1.92) and pre-therapy CORE-OM non-risk scores did not demonstrate a curvilinear relationship with the outcome.

To aid interpretation, graphical representations of the number of patients who met criteria for RCSI were produced for each significant predictor. Pre-therapy CORE-OM non-risk scores were grouped to produce categories of scores which ranged in 5 points. Figure 3 indicates that RCSI rates were similar across patients with pre-therapy CORE-OM non-risk scores between 10 and 20. As pre-therapy CORE-OM non-risk scores increased from 20, the rate of RCSI steadily fell.

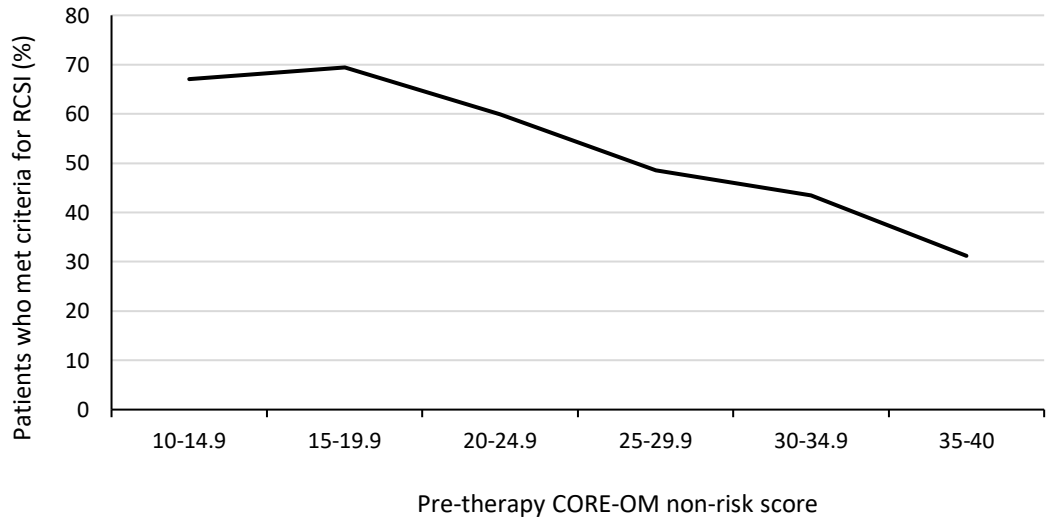

Figure 3. Line graph depicting rates of reliable and clinically significant improvement (RCSI) across grouped pre-therapy Clinical Outcomes in Routine Evaluation-Outcome Measure (CORE-OM) non-risk scores.

Figure 4 shows the percentages of patients who met criteria for RCSI across categories of medication use. RCSI rates were highest for patients with medication reductions, and lowest for patients with medication increases/modifications.

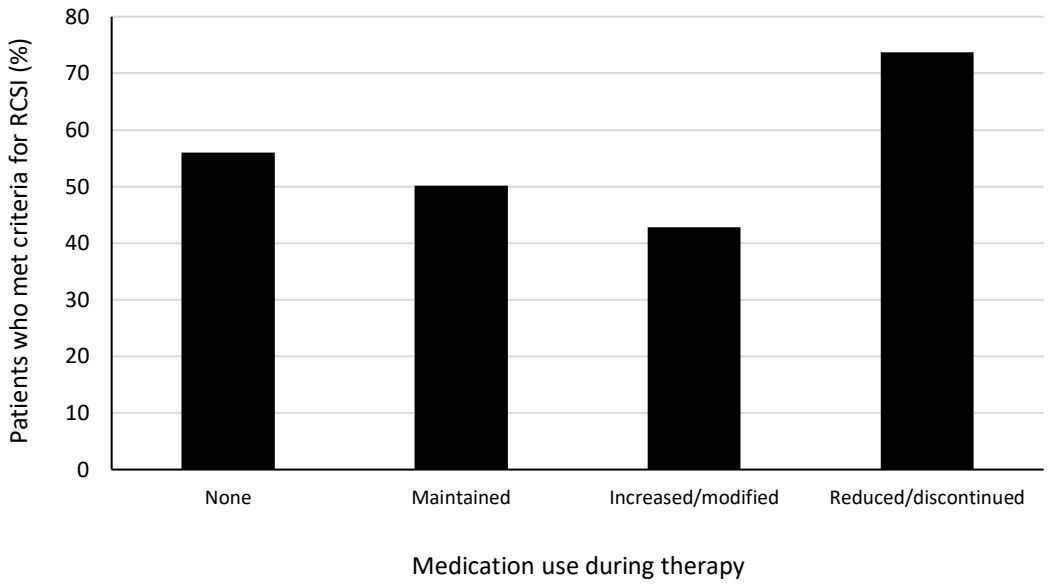

Figure 4. Bar chart depicting rates of reliable and clinically significant improvement (RCSI) across categories of medication use during therapy.

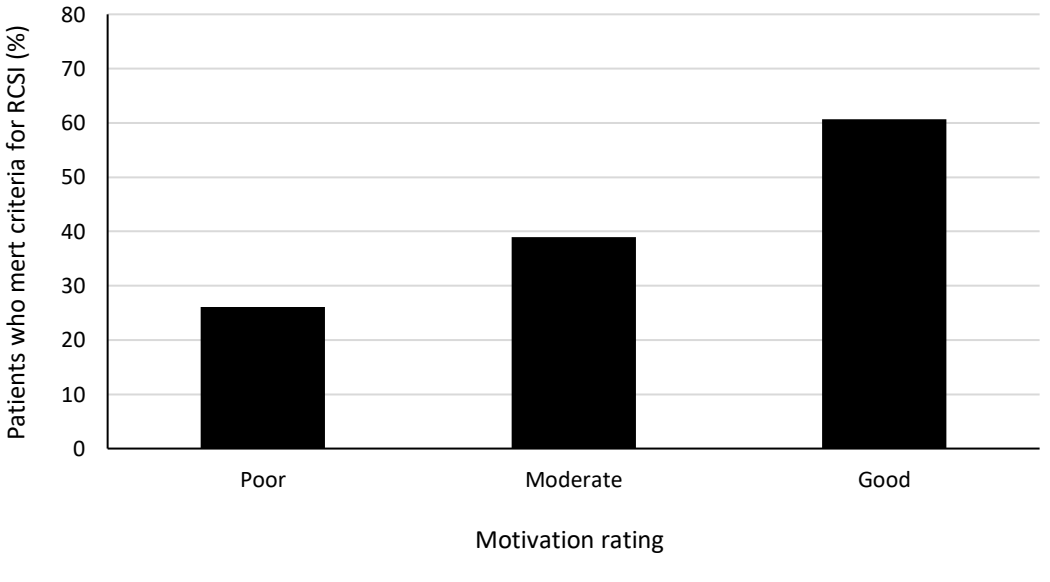

Figure 5. Bar chart depicting rates of reliable and clinically significant improvement (RCSI) across categories of motivation ratings.

Figure 5 shows the percentages of patients who met criteria for RCSI across

each motivation rating. RCSI rates increased across the categories of poor, moderate, and good motivation. The RCSI rate increased more steeply between the categories of moderate and good motivation compared to between the categories of poor and moderate motivation.

Figure 6 shows the percentages of patients who met criteria for RCSI for each category of employment status. A larger percentage of patients who were employed or in education met criteria for RCSI compared to patients who were unemployed.

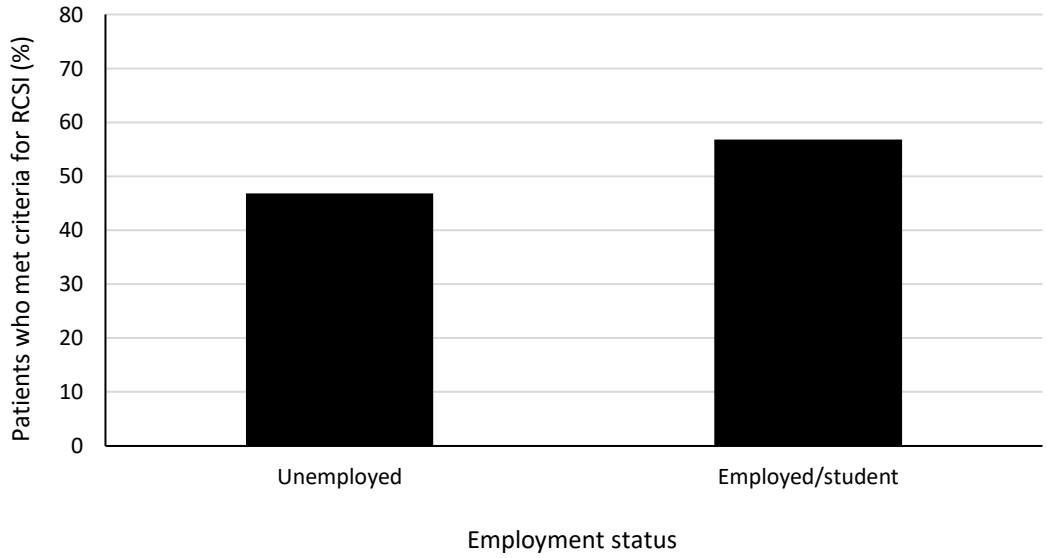

Figure 6. Bar chart depicting rates of reliable and clinically significant improvement (RCSI) across categories of employment status.

Therapist effect analyses. Table 8 shows the therapist effect estimates for both the MQL- and PQL-developed models. Wald tests indicated that there was a significant portion of unexplained variance between therapists regarding the likelihood of patients meeting criteria for RCSI in both the MQL-developed model, $\chi^2(1) = 28.23$, $p < .001$, and the PQL-developed model, $\chi^2(1) = 29.06$, $p < .001$. After accounting for patient-level predictors of outcome, estimates suggested that between 11.3% and 13.0% of the variance in outcomes was attributable to therapist variability, depending on the approximation procedure used.

Table 8.

Estimates of Therapist-Level Variance, VPC Statistics, and Therapist Effects

Approximation procedure	σ^2_{u0}	VPC	Therapist effect (%)
MQL	0.418	0.11273	11.27
PQL	0.491	0.12986	12.99

Note. σ^2_{u0} = therapist-level variance. VPC = variance partition coefficient. MQL = marginal quasi-likelihood. PQL = penalised quasi-likelihood.

Figure 7 depicts the therapist residuals with 99.9% confidence intervals. Therapists were ranked in order of the extent to which their outcomes differed from the mean odds of having a patient meet criteria for RCSI. Negative residuals represent therapists with less patients meeting criteria for RCSI than average, and positive residuals represent therapists with more patients meeting criteria for RCSI than average. Where the confidence intervals for therapist residuals do not cross zero, those therapists were significantly more or less likely to have patients meet criteria for RCSI.

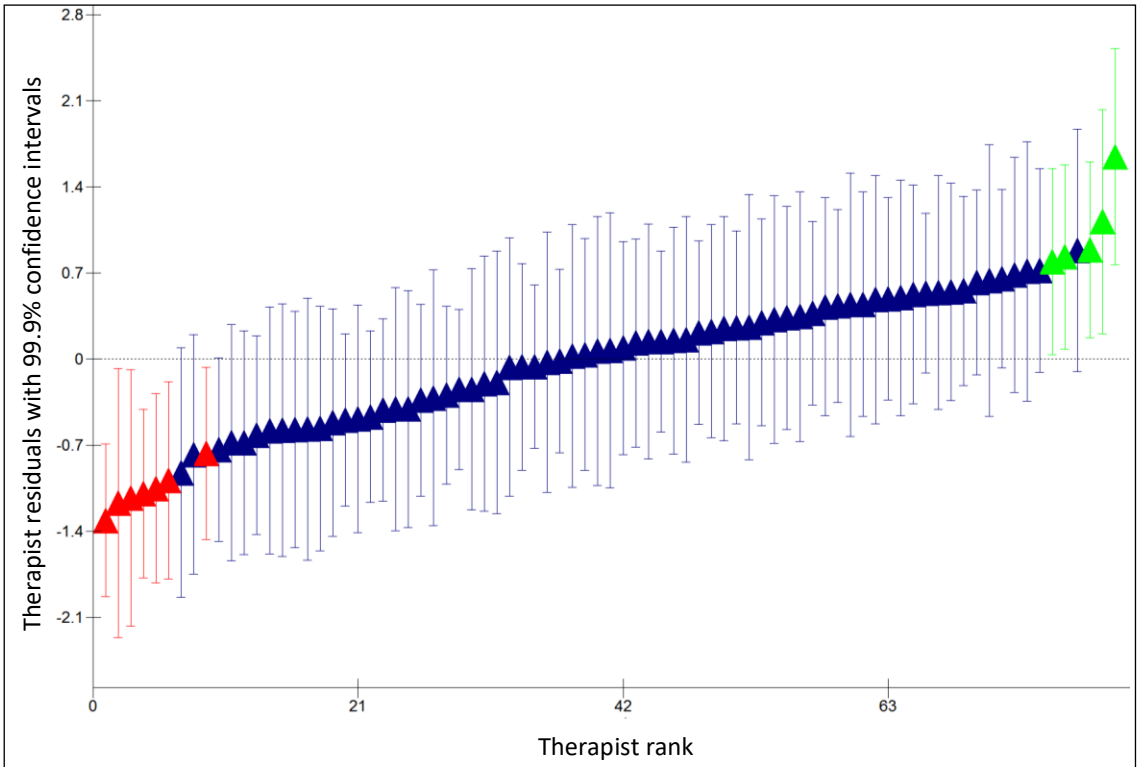

Figure 7. Therapist residuals with 99.9% confidence intervals. ▲ = Residuals for therapists significantly less likely to have patients meet criteria for reliable and clinically significant improvement than average. ▲ = Residuals for therapists significantly more likely to have patients meet criteria for reliable and clinically significant improvement than average.

Figure 7 shows that 7 therapists (8.64%) were significantly less likely to have patients meet criteria for RCSI than average and 5 therapists (6.17%) were significantly more likely to have patients meet criteria for RCSI than average. When therapist residuals with 95% confidence intervals were examined, 16 therapists (19.75%) were significantly less likely to have patients meet criteria for RCSI than average and 14 therapists (17.28%) were significantly more likely to have patients meet criteria for RCSI than average.

Table 9 presents descriptive statistics for the RCSI rates of the below average, average, and above average therapist groups. As Table 9 shows, the mean RCSI rate among below average therapists was 26.41% lower than the mean RCSI rate across average therapists. The mean RCSI rate among above average therapists was 22.88% higher than the mean RCSI rate across average therapists. The 95% confidence intervals for mean RCSI rates across each therapist group did not overlap.

Table 9.

Descriptive Statistics for RCSI Rates Among Below Average, Average, and Above Average Therapists

			RCSI rate			
Therapist group	Therapist *n*	Patient *n*	*M*	*SD*	95% CIs	Range
Below average	7	571	29.35	7.21	[24.00, 34.69]	20.83–39.62
Average	69	4,023	55.76	12.91	[52.71, 58.81]	27.78–77.42
Above average	5	382	78.64	5.94	[73.44, 83.85]	70.18–85.96

Note. CIs = confidence intervals.

Therapist residuals were plotted against therapist RCSI rates (see Figure 8). Figure 8 indicates a linear trend of increasing RCSI rate. The most effective therapist appeared to be an outlier, with a much larger residual compared to other therapists.

Multilevel modelling assumptions. Plots of standardised residuals against normally distributed scores indicated that the assumption of normality had been met. To explore the assumption of homoscedasticity, standardised residuals were plotted against

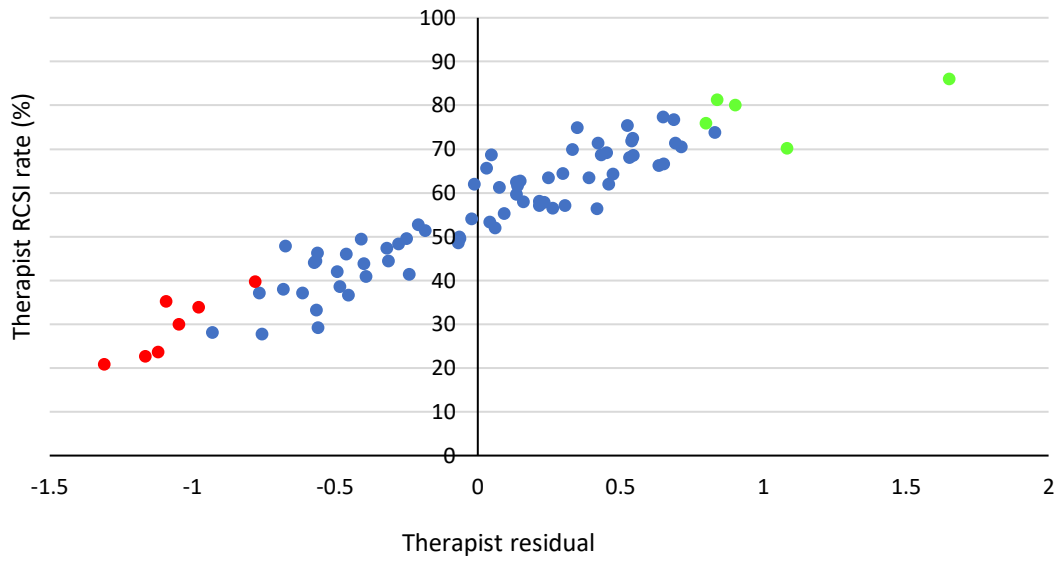

Figure 8. Scatterplot of therapist residuals against therapist rates of reliable and clinically significant improvement (RCSI). ● = Data for therapists who were significantly less effective than average. ● = Data for therapists who were significantly more effective than average.

predicted scores at level two of the data. The data appeared relatively homogeneous, except for one outlier. The final model was re-run without data from this therapist using MQL approximation procedures (see Appendix M). The results of this analysis were largely equivalent with the original results, except for the estimated therapist effect which reduced from 11.3% to 10.3%. No data entry errors were identified. Thus, data from this therapist remained in the main analysis to represent a true picture of all available data.

Sensitivity analyses. To test the robustness of the final model, the multilevel modelling analysis was repeated using data from all therapists (n = 1,020) regardless of the number of patients they submitted data for (patient n = 8,885). This model is presented in Table 10 (see Appendix N for the full model). The results of this analysis were largely equivalent with the original results, except that the categories of *poor* and *moderate* motivation made significantly different contributions to the model, $\chi^2(1) = 17.73$, $p < .001$. No significant random slopes were found ($ps \geq 0.33$) and no interaction effects were found among predictors ($zs \leq 2.10$).

Table 10.

Two-Level Model of Patient-Level Predictors and Therapist Effects in Predicting the Likelihood of Meeting Criteria for RCSI Across the Whole Therapist Sample

Variable	β	*SE*	*z*	*OR*	95% CIs for OR
MQL estimates					
Pre-therapy CORE-OM non-risk score	-0.07	0.01	-14.60	0.93	[0.91, 0.95]
Reduced medication[a]	0.80	0.10	8.35	2.23	[1.83, 2.71]
Increased/maintained medication[a]	-0.14	0.05	-2.82	0.87	[0.79, 0.96]
Poor motivation[b]	-1.47	0.13	-11.69	0.23	[0.18, 0.30]
Moderate motivation[b]	-0.92	0.06	-16.43	0.40	[0.35, 0.45]
Unemployed[c]	-0.44	0.06	-7.33	0.64	[0.57, 0.72]
PQL estimates					
Pre-therapy CORE-OM non-risk score	-0.08	0.01	-15.80	0.92	[0.91, 0.94]
Reduced medication[a]	0.88	0.10	8.76	2.41	[1.98, 2.93]
Increased/maintained medication[a]	-0.16	0.05	-3.00	0.85	[0.77, 0.94]
Poor motivation[b]	-1.62	0.13	-12.46	0.20	[0.15, 0.26]
Moderate motivation[b]	-1.01	0.06	-17.41	0.36	[0.32, 0.41]
Unemployed[c]	-0.48	0.06	-7.67	0.62	[0.55, 0.70]

Note. OR = odds ratio. CI = confidence interval. MQL = marginal quasi-likelihood. CORE-OM = Clinical Outcomes in Routine Evaluation-Outcome Measure. PQL = penalised quasi-likelihood.

[a] reference category *no medication*. [b] reference category *good* motivation. [c] reference category *employed/student*.

Table 11.

Estimates of Therapist-Level Variance, VPC Statistics, and Therapist Effects Using the Whole Therapist Sample

Approximation procedure	σ^2_{u0}	VPC	Therapist effect (%)
MQL	0.338	0.09316	9.32
PQL	0.409	0.11057	11.06

Note. σ^2_{u0} = therapist-level variance. VPC = variance partition coefficient. MQL = marginal quasi-likelihood. PQL = penalised quasi-likelihood.

Table 11 shows the therapist effect estimates for both the MQL- and PQL-developed models using the whole therapist sample. Wald tests indicated that there was a significant portion of unexplained variance between therapists with regards to the likelihood of patients experiencing RCSI in both the MQL-developed model, $\chi^2(1) = 54.86$, $p < .001$, and the PQL-developed model, $\chi^2(1) = 61.44$, $p < .001$. Therapist effect estimates ranged from 9.3% to 11.1%, depending on the approximation procedure used. These estimates were slightly smaller than those obtained in the analysis of therapists

who provided data for ≥ 30 patients.

Discussion

This study had three main aims. First, it aimed to investigate variability in risk-to-self across patients accessing different service types. Second, it aimed to investigate rates of reliable and clinically significant change in therapy among at-risk patients. Third, it aimed to identify predictors of meeting criteria for RCSI in therapy among at-risk patients. The results are discussed below in relation to each aim.

Variability in Risk and Non-Risk Scores Across Services

Risk-to-self scores were compared between patients accessing NHS primary care, voluntary, and university/workplace services. The results indicated that patients in university/workplace services reported significantly lower risk-to-self than patients in primary care or voluntary services. Accordingly, a previous practice-based study using a large dataset also found that patients in university services reported significantly lower risk than patients in primary care (Connell, Barkham, & Mellor-Clark, 2007). Regarding CORE-OM non-risk scores, patients in university/workplace services had significantly higher scores than patients in voluntary services and significantly lower scores than patients in primary care. These results suggest that, irrespective of general distress, patients in university/workplace services present with lower risk-to-self than patients who access voluntary or primary care services. However, it must be noted that substantially more patients accessed university/workplace services than other types of services, which may have impacted the statistical power of analyses. Indeed, descriptive statistics indicated that patients in private services had lower risk-to-self scores (M = 3.79) than patients in university/workplace services (M = 3.97), but no significant differences were found between risk-to-self scores in private services and scores in any sector other than secondary care. Therefore, further large-scale replications of this finding are needed to support its validity.

Among NHS primary, secondary, and tertiary care services, risk-to-self scores were expected to be lowest among primary care patients and highest among tertiary care patients, in line with the stepped care model. Contrary to this hypothesis, secondary care patients reported significantly higher risk-to-self than patients in any other sector, and scores did not significantly differ between patients in primary and tertiary care. Previous research supports that secondary care patients report significantly greater risk than primary care patients (Barkham et al., 2005). It was interesting that risk-to-self scores did not differ between tertiary care patients and patients from any other service type, given that tertiary care includes specialist services for patients considered at-risk of harming themselves or others (Bowers et al., 2005). The proportion of patients accessing tertiary care was very small (0.39%), which may have reduced statistical power to detect between-group differences. Alternatively, tertiary care provisions may offer protective factors against risk-to-self, such as increased social support (from other patients and staff) and reduced access to means of self-harm (Hunt et al., 2014; James, Stewart, & Bowers, 2012). Indeed, by the nature of accessing tertiary care services, patients typically receive multidisciplinary interventions before, or alongside, psychological therapy. This may account for the lower risk-to-self among tertiary care patients at entry to therapy.

Reliable and Clinically Significant Change Among At-Risk Patients

Among at-risk patients, pre-post therapy CORE-OM scores reduced with a large effect size. The non-pooled effect size was $d = 2.0$ and the effect size calculated using Morris and DeShon's (2002) correction for dependency was $d = 1.5$. These estimates are high in relation to those reported in previous practice-based studies of patients who complete therapy. Previous studies have reported effect sizes of 1.6 on the CORE-OM (Saxon & Barkham, 2012) and 1.4 on the Patient Health Questionnaire (Richards & Suckling, 2009). Indeed, a meta-analysis of ten-practice based studies of therapy for

common mental health disorders reported an uncontrolled effect size of 1.29 (Cahill, Barkham, & Stiles, 2010). This suggests that therapy is at least equally effective in reducing distress among at-risk patients in comparison to studies of general samples.

Regarding reliable and clinically significant change, 51.33% of at-risk patients met criteria for RCSI in therapy. Compared to a normal distribution, patients were more likely to meet criteria for clinically significant and/or reliable improvement and less likely to meet criteria for clinically significant and/or reliable deterioration. The RCSI rate was lower than that reported by a previous study which sampled data from the same original dataset as the current study. Stiles et al. (2015) reported an RCSI rate of 60% among patients with planned therapy endings. The current study used a reliable change index of 5 points, which is commonly cited for the CORE-OM (Barkham et al., 2006), whereas Stiles et al. (2015) used a study-specific reliable change index of 4.5. Therefore, to meet criteria for RCSI, patients in the current study required slightly greater changes in CORE-OM scores. Nonetheless, a higher proportion of the current sample met criteria for reliable improvement compared to Stiles et al.'s (2015) sample (27.1% versus 19.9%, respectively). Furthermore, Stiles et al.'s (2015) result was supported by Saxon and Barkham's (2012) practice-based study of patients across primary care and counselling services, which reported a similar RCSI rate of 61.6%. Therefore, the RCSI rate may have been lower in the current study due to its focus on at-risk patients, thereby suggesting that therapy facilitates less RCSI among at-risk patients compared to general samples.

Predictors of RCSI for At-Risk Patients

A substantial number of variables were assessed to identify differences between patients who did and did not meet criteria for RCSI. Multilevel models were developed to assess the contribution of patient-related variables and therapist effects in predicting the likelihood of patients meeting criteria for RCSI. Patients with lower pre-therapy

CORE-OM non-risk scores, higher motivation, reduced medication, and those in employment or education were significantly more likely to meet criteria for RCSI. Accounting for these factors, a significant portion of the variance in outcomes was attributable to therapist effects. Each of these predictors is discussed below.

Pre-therapy CORE-OM non-risk scores. Patients with higher pre-therapy CORE-OM non-risk scores (i.e., higher initial distress) were less likely to meet criteria for RCSI. This corresponds with findings that higher pre-therapy symptom severity is associated with poorer outcomes across a range of clinical presentations and therapeutic modalities (Cooper et al., 2016; Firth et al., 2015; Knopp et al., 2013; Saxon & Barkham, 2012). However, previous studies with at-risk patients have reported inconsistent findings regarding the relationship between pre-therapy symptom severity and outcomes. Whilst higher pre-therapy symptom severity was associated with poorer outcomes in Guthrie et al.'s (2003) study of patients who presented to hospital after self-poisoning, it was associated with better outcomes in Gratz et al.'s (2014) study of patients with borderline personality disorder symptoms. These studies included small and relatively homogeneous samples. Using a large and diverse sample, the current study supports the conclusion that higher pre-therapy distress predicts poorer outcomes among at-risk patients. However, it must be noted that only 2.35% of the sample with which the multilevel model was developed reported personality-related difficulties. Given other findings from small studies that higher psychopathology is associated with better outcomes among patients with borderline personality disorder diagnoses (Barnicot et al., 2012), future large-scale investigations could clarify whether pre-therapy distress relates differently to outcomes among this population.

Medication. Patients whose psychiatric medication was reduced or discontinued during therapy were more likely to meet criteria for RCSI. Unfortunately, the timing of medication reductions was not recorded. Given that one CORE-OM subscale measures

symptoms, it may be that both RCSI outcomes and medication reductions reflected symptom improvement. Alternatively, medication reductions may provide specific recovery-enhancing benefits. Medication reductions could be experienced as an objective sign of improvement (i.e., from the prescriber) that patients with no medication or medication increases do not have the opportunity to experience. A systematic review of studies of routine outcome measuring indicated that therapy outcomes improve when patients receive feedback about their progress (Gondek, Edbrooke-Childs, Fink, Deighton, & Wolpert, 2016). Such feedback may impact patients' hope, self-efficacy, or motivation, all of which are associated with better outcomes (Brown et al., 2014; Romano & Peters, 2015). However, medication increases, which could be interpreted as negative feedback, did not significantly contribute to the model. A further tentative hypothesis is that medication reductions may be associated with improvements in functioning. In a longitudinal randomised study of patients with psychosis, medication reduction regimens were associated with twice the recovery rate as medication maintenance regimens (Wunderlink, Nieboer, Wiersma, Sytema, & Nienhuis, 2013). This result appeared to relate to higher functioning across self-care, interpersonal, and occupational domains, rather than symptom remission, which the authors discussed in relation to the reduced impact of medication on executive functioning, drive, and alertness. Indeed, one of the CORE-OM subscales specifically measures patient functioning across similar domains.

Motivation. Patients with poor or moderate motivation were less likely to meet criteria for RCSI than patients with good motivation. It must be acknowledged that ratings of patient motivation were provided by therapists at the end of therapy. This enhanced the risk of bias, as therapists may have rated patients who they perceived to have made less progress as being lower in motivation. Nonetheless, previous studies support a positive relationship between motivation and therapy outcomes across various

clinical populations (Langner et al., 2009; Romano & Peters, 2015; Vall & Wade, 2015). Self-determination theory suggests that motivation levels may be affected by perceived levels of autonomy, positive feedback, and interpersonal support (Deci & Ryan, 1985). In line with this theory, there are several tentative hypotheses for the relationship between motivation and outcomes. Firstly, greater motivation may be associated with increased self-efficacy, which itself is predictive of better therapy outcomes (Brown et al., 2014). Secondly, effective therapies may inherently promote patient motivation by providing positive feedback (e.g., symptom reduction). Thirdly, motivation may be associated with patient perceptions of therapist support. This study found a large-sized positive relationship between ratings of patients' motivation and working alliances. Consequently, working alliance was not entered into the model to safeguard against multicollinearity. Given that better quality working alliances are associated with better therapy outcomes (Dinger, Strack, Leichsenring, Wilmers, & Schauenberg, 2008), the relationship between motivation and outcomes may be associated with patients' level of support.

Employment. Patients who were unemployed were less likely to meet criteria for RCSI than patients who were employed or in education. Interestingly, employment status only significantly contributed to the model after therapist variability was taken into account. As 41.86% of the sample were accessing university/workplace services, it may be that the effect of employment status was suppressed prior to controlling for therapist variability. Previous studies support the relationship between unemployment and poorer therapy outcomes (Firth et al., 2015; Frank et al., 2002). Several hypotheses could account for this relationship. In Milner, Page, and LaMontagne's (2014) random-effects meta-analysis, unemployment was associated with a significantly higher relative risk of suicide and mental health difficulties over time. Thus, unemployment may be associated with increased psychopathology which, as discussed above, is associated

with poorer outcomes. Alternatively, employment may offer protective factors which promote opportunities for therapeutic change. In a qualitative study with 23 participants who identified as having recovered from mental health difficulties, participants reported that employment had promoted their pride, self-esteem, financial security, and coping resources which, ultimately, facilitated their mental health recovery (Dunn, Wewiorski, & Rogers, 2008).

Therapist effects. It was hypothesised that therapist effects would account for up to 10% of the variability in therapy outcomes. The results indicated that between 10.3% and 13.0% of the variance in outcomes was attributable to therapist effects, depending on the approximation procedure used and whether data from an outlier were included. These estimates are high in comparison to those reported by previous studies which have modelled therapist effects and patient-related variables. Previous therapist effect estimates have ranged from 5% to 12% (Green et al., 2014; Kim, Wampold, & Bolt, 2006; Saxon, Firth, & Barkham, 2016). As in previous studies (Okiishi, Lambert, Nielsen, & Ogles, 2003; Saxon & Barkham, 2012), the results suggested that some therapists were significantly more effective than average and some were significantly less effective than average.

There are several potential explanations for the high therapist effect. Previous studies have found that therapist effect estimates reduce when therapist-level variables (e.g., caseload mix) are added to models (Firth et al., 2015; Saxon & Barkham, 2012). This suggests that some of the therapist effect may have been attributable to unmodeled therapist variables. A further tentative hypothesis for the high therapist effect regards therapeutic provisions. Many previous studies of therapist effects have included therapists who provided manualised interventions (Firth et al., 2015; Green et al., 2014; Kim et al., 2006), which are associated with reduced therapist variability (Crits-Cristoph et al., 1991). In contrast, 69.18% of the sample with which the current multilevel model

was developed received integrative therapies. Furthermore, the present study was the first to include only at-risk patients. Whilst risk scores did not significantly predict outcomes, previous findings indicate that larger therapist risk caseloads are associated with poorer outcomes, and therapist effect estimates reduce when therapist risk caseloads are accounted for (Saxon & Barkham, 2012). Accordingly, in the current study, the therapist effect estimate was lower in the analysis of the whole therapist sample compared to the analysis of therapists who provided data for ≥ 30 patients. Thus, the higher therapist effect may have been attributable to the analysis of therapists with substantial risk caseloads.

Critique

The current study has several strengths. The use of a national practice-based dataset enabled the inclusion of a large sample with various difficulties and comorbidities that presented for routine therapy across a range of services. This may have enhanced both the generalisability and ecological validity of the results. Additionally, the large sample size permitted the use of multilevel modelling, which enabled the modelling of therapist variability alongside patient-level predictors of outcomes. This method may produce more realistic estimates of treatment effects compared to methods which do not model therapist variability (Wampold & Serlin, 2000). Furthermore, many potential patient-level predictors were assessed, which permitted a comprehensive evaluation of variables implicated in RCSI among at-risk patients. The use of a conservative alpha level (< .001) throughout analyses, and the computation of effect sizes to inform modelling decisions, may have enhanced the reliability of results. The reliability of results was further strengthened by the inclusion of a large therapist sample with each therapist providing data for ≥ 30 patients in the multilevel modelling analysis, and the confirmation of this analysis's results with a larger therapist sample. Finally, therapy outcomes were defined in terms of reliable and

clinically significant change (Jacobson & Truax, 1991), which took into account potential measurement error and whether outcomes indicated changes in patients' clinical presentations.

Several limitations should also be considered when interpreting this study's results. It must be acknowledged that the study-specific dataset was derived from a much larger dataset. A reduction in sample size occurred to meet multilevel modelling requirements and the study's sampling criteria. In line with the sampling criteria, the results may not be representative of patients with moderate-severe cognitive difficulties and patients accessing non-individual therapies (e.g., group therapy). These exclusions were undertaken as the CORE-OM is not validated for use with people with intellectual disabilities, and information was not available to model group therapy effects. Further patients were excluded due to having incomplete outcome measures. Compared to intent-to-treat analyses, analysing data for only patients who complete therapy may overestimate treatment effects. To counterbalance this, patients' data were analysed regardless of their attendance or whether they had unplanned therapy endings. Nonetheless, included patients had significantly higher attendance rates and planned endings than excluded patients, and research suggests that therapist effects are larger when patients who drop out of therapy are included in analyses (Lutz, Leon, Martinovich, Lyons, & Stiles, 2007).

Further limitations may apply to the study measures. The use of a single self-report outcome measure has drawbacks in having a narrow focus and being at risk of response bias. However, the CORE-OM is a well validated measure of distress across several domains, and its convergent validity with other practitioner- and patient-completed measures has been supported (Barkham et al., 2001; Evans et al., 2002). Additionally, patients are arguably best placed to provide measures of their own distress. Another limitation was the use of the CORE-OM risk-to-self items to classify

patients as “at-risk”. The sample is likely to have included patients with diverse risk-related difficulties, ranging from thoughts of self-harm to suicide attempts. Given arguments that suicidal and non-suicidal self-injury represent distinct conditions (Nock, 2009), this complicated the ability to draw conclusions for specific populations. However, as CORE-OM risk scores correlate positively with therapists’ clinical risk assessments (Bedford, Lukic, & Tibbles, 2011; Whewell & Bonanno, 2000), they were considered an appropriate proxy measure in the absence of information regarding patients’ actual self-injury. Finally, the use of therapist-completed measures to identify predictor variables had limitations. Given findings that therapist and patient ratings of therapy elements (e.g., working alliances) often differ (Tyron, Blackwell, & Hammel, 2007), some of the study variables may have reflected subjective therapist opinions.

Conclusions and Implications for Practice and Research

In conclusion, both patient-related variables and therapist effects significantly impact the likelihood of at-risk patients achieving RCSI in therapy. Across a large sample of at-risk patients, 51.33% achieved RCSI in therapy and a further 27.12% reliably improved. This suggests that at-risk patients generally experience a good enough service, but there is room for improvement. Whilst the cross-sectional design of this study impedes the ability to infer causal relationships, several directions for clinical practice and research may be deduced.

Regarding patient-related predictors of outcomes, the results suggest that clinical efforts to enhance patient motivation (see Ryan, Lynch, Vansteenkiste, & Deci, 2011) may promote the likelihood of patients achieving RCSI. Motivation is a multifaceted concept (Deci & Ryan, 1985), and future studies could disentangle the effects of intrinsic and extrinsic motivation on outcomes. This could enable therapists to promote the type of motivation most strongly linked to recovery.

The relationship between reduced medication and RCSI was interesting. Whilst

a direct relationship between reducing medication and achieving RCSI should not be inferred, clinicians are advised to monitor outcomes among patients whose medications are maintained or increased during therapy to identify lack of change. Longitudinal investigations of the direction of the relationship between medication reductions and RCSI could determine whether these factors represent a similar underlying phenomenon (e.g., symptom reduction) or whether medication reductions promote recovery for at-risk patients.

The results also suggest that efforts to promote at-risk patients' access to employment and education may increase their likelihood of achieving RCSI in therapy. In the context of the integration of health and social services in the United Kingdom and the pooling of their resources (Department of Health, 2017), there may be opportunities for services to work jointly on projects which enhance employment opportunities for at-risk patients. At the individual level, identifying occupational goals with patients and incorporating these into multidisciplinary careplans may be beneficial.

The final patient-level predictor of achieving RCSI was pre-therapy distress. This finding indicates that therapy may be most effective in facilitating recovery when patients present with less distress, which lends further support to calls for the promotion of early intervention in mental health services (Mental Health Taskforce, 2016). The contribution of pre-therapy distress to the model was relatively unchanged when other variables were added, thus the factors implicated in the relationship between pre-therapy distress and RCSI were unclear. This finding has been repeatedly reported and clearly patients with higher pre-therapy distress need something different in therapy to recover. Future large-scale, practice-based studies could focus on patients with high pre-therapy distress to determine predictors of outcomes for this population.

After patient-related variables were taken into account, 10-13% of the variance in outcomes was attributable to therapist effects. Some therapists were significantly

more effective, and some were significantly less effective, than average. As therapists were treated as a random sample, these results can be broadly generalised. The high therapist effect in comparison to studies which included patients not at-risk suggests the need for careful consideration around allocating at-risk patients to therapists (see also Saxon & Barkham, 2012). The current findings support the continued use of routine outcome measuring of therapist caseloads, which could inform supervisory processes, evaluation, and service development. Identifying effective therapists could provide opportunities for shared learning, and identifying less effective therapists could enable their targeted support in terms of supervision and case management. Importantly, previous studies have indicated that therapist-related variables such as type and amount of training are not predictive of therapy outcomes (Okiishi et al., 2006). It would be useful for future studies to identify therapist-level variables associated with outcomes, to identify potential avenues for improving therapists' effectiveness. Unfortunately, analyses of therapist effects have been largely limited to studies using national datasets to meet sample size requirements, and such datasets typically include little information about therapists. This has implications for the development of large datasets which record more information about therapists.

This was the first study to assess the relative impact of patient-level variables and therapist effects in predicting RCSI among at-risk patients in therapy. Future studies should aim to replicate the model using other datasets and different outcome measures. Like many previous studies of therapist effects (Green et al., 2014; Saxon & Barkham, 2012), most of the current patients accessed primary care and university/workplace services. Consequently, the level of risk-to-self in the current sample may be low compared to patients accessing specialist services. This study could not model service-level variation in outcomes as few data were available from services at higher tiers of the stepped care model. Secondary and tertiary care providers are encouraged to submit

data to national datasets or create unique datasets for their services, to enable thorough practice-based evaluations of outcomes for patients accessing these services.

Furthermore, this study focused on RCSI outcomes. Whilst only 3% of patients deteriorated, 20.25% made no reliable change. Future studies could identify predictors of lack of improvement among patients, to further inform the development of effective therapies. Finally, this study adds to the growing literature which suggests that therapists significantly vary in effectiveness. This has crucial implications for the measurement of therapist effects in therapy outcome research, to provide accurate estimates of therapy effectiveness.

References

Barkham, M., Evans, C., Margison, F., McGrath, G., Mellor-Clark, J., Milne, D., & Connell, J. (1998). The rationale for developing and implementing core outcome batteries for routine use in service settings and psychotherapy outcome research. *Journal of Mental Health, 7,* 35-47. doi:10.1080/09638239818328

Barkham, M., Gilbert, N., Connell, J., Marshall, C., & Twigg, E. (2005). Suitability and utility of the CORE-OM and CORE-A for assessing severity of presenting problems in psychological therapy services based in primary and secondary care settings. *British Journal of Psychiatry, 186,* 239-246. doi:10.1192/bjp.186.3.239

Barkham, M., Hardy, G., & Mellor-Clark, J. (2010). *Developing and delivering practice-based evidence: A guide for the psychological therapies.* Chichester, UK: Wiley.

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

Barkham, M., & Mellor-Clark, J. (2003). Bridging evidence-based practice and practice based evidence: Developing a rigorous and relevant knowledge for the psychological therapies. *Clinical Psychology & Psychotherapy, 10,* 319-327. doi:10.1002/cpp.379

Barkham, M., Mellor-Clark, J., Connell, J., & Cahill, J. (2006). A core approach to practice-based evidence: A brief history of the origins and applications of the CORE-OM and CORE system. *Counselling and Psychotherapy Research, 6,*

3-15. doi:10.1080/14733140600581218

Barnicot, K., Katsakou, C., Bhatti, N., Savill, M., Fearns, N., & Pribe, S. (2012). Factors predicting the outcome of psychotherapy for borderline personality disorder: A systematic review. *Clinical Psychology Review, 32,* 400-412. doi:10.1016/j.cpr.2012.04.004

Bedford, A., Lukic, G., & Tibbles, J. (2011). Evaluation of risk by patients' and with clinicians' ratings: A CORE-OM and CORE-A investigation. *Clinical Psychology and Psychotherapy, 18,* 244-249. doi:10.1002/cpp.714

Bedics, J. D., Atkins, D. C., Harned, M. S., & Linehan, M. M. (2015). The therapeutic alliance as a predictor of outcome in dialectical behavior therapy versus nonbehavioral psychotherapy by experts for borderline personality disorder. *Psychotherapy, 52,* 67-77. doi:10.1037/a0038457

Beutler, L. E., Malik, M. L., Alimohamed, S., Harwood, T. M., Talebi, H., & Noble, S. (2003). Therapist variables. In M. J. Lambert (Ed.), *Handbook of psychotherapy and behavior change* (5th ed., pp. 227-306). New York, NY: Wiley.

Bower, P., & Gilbody, S. (2005). Stepped care in psychological therapies: Access, effectiveness and efficiency. Narrative literature review. *British Journal of Psychiatry, 186,* 11-17. doi:10.1192/bjp.186.1.11

Bowers, L., Simpson, A., Alexander, J., Hackney, D., Nijman, H., Grange, A., & Warren, J. (2005). The nature and purpose of acute psychiatric wards: The Tompkins acute ward study. *Journal of Mental Health, 14,* 625-635. doi:10.1080/09638230500389105

Brown, L. A., Wiley, J. F., Wolitzky-Taylor, K., Roy-Byrne, P., Sherbourne, C., Stein, M. B., … Craske, M. G. (2014). Changes in self-efficacy and outcome expectancy as predictors of anxiety outcomes from the CALM study.

Depression and Anxiety, 31, 678-689. doi:10.1002/da.22256

Cahill, J., Barkham, M., & Stiles, W. B. (2010). Systematic review of practice-based research on psychological therapies in routine clinic settings. *British Journal of Clinical Psychology, 49,* 421-453. doi:10.1348/014466509X470789

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Earlbaum.

Comtois, K. A., & Linehan, M. M. (2006). Psychosocial treatments of suicidal behaviors: A practice-friendly review. *Journal of Clinical Psychology, 62,* 161-170. doi:10.1002/jclp.20220

Connell, J., Barkham, M., & Mellor-Clark, J. (2007). CORE-OM mental health norms of students attending university counselling services benchmarked against an age-matched primary care sample. *British Journal of Guidance & Counselling, 35,* 41-57. doi:10.1080/03069880601106781

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

Cooper, Z., Allen, E., Bailey-Straebler, S., Basden, S., Murphy, R., O'Connor, M. E., & Fairburn, C. G. (2016). Predictors and moderators of response to enhanced cognitive behaviour therapy and interpersonal psychotherapy for the treatment of eating disorders. *Behaviour Research and Therapy, 84,* 9-13. doi:10.1016/j.brat.2016.07.002

Crawford, M. J., Thomas, O., Khan, N., & Kulinskaya, E. (2007). Psychosocial interventions following self-harm: Systematic review of their efficacy in preventing suicide. *British Journal of Psychiatry, 190,* 11-17. doi:10.1192/bjp.bp.106.025437

Crits-Cristoph, P., Baranackie, K., Kurcias, J., Beck, A., Carroll, K., Perry, K., … Zitrin, C. (1991). Meta-analysis of therapist effects in psychotherapy outcome studies. *Psychotherapy Research, 1,* 81-91. doi:10.1080/10503309112331335511

Crits-Cristoph, P., Tu, X., & Gallop, R. (2003). Therapists as fixed versus random effects-some statistical and conceptual issues: A comment on Siemer and Joormann (2003). *Psychological Methods, 8,* 518-523. doi:10.1037/1082-989X.8.4.518

Daigle, M. S., Pouliot, L., Chagnon, F., Greenfield, B., & Mishara, B. (2011). Suicide attempts: Prevention of repetition. *Canadian Journal of Psychiatry, 56,* 621-629. Retrieved from http://journals.sagepub.com/doi/pdf/10.1177/070674371105601008

Davidson, K., Scott, J., Schmidt, U., Tata, P., Thornton, S., & Tyrer, P. (2004). Therapist competence and clinical outcome in the Prevention of Parasuicide by Manual Assisted Cognitive Behaviour Therapy Trial: The POPMACT study. *Psychological Medicine, 34,* 855-863. doi:10.1017/S0033291703001855

Deci, E. L., & Ryan, R. M. (1985). *Intrinsic motivation and self-determination in human behavior.* New York, NY: Plenum.

Department of Health. (2017). *Integration and better care fund policy framework 2017-2019.* London, UK: Author.

Dinger, U., Strack, M., Leichsenring, F., Wilmers, F., & Schauenburg, H. (2008). Therapist effects on outcome and alliance in inpatient psychotherapy. *Journal of Clinical Psychology, 64,* 344-354. doi:10.1002/jclp.20443

Drummond, C., Coulton, S., James, D., Godfrey, C., Parrott, S., Baxter, J., … Peters, T. (2009). Effectiveness and cost-effectiveness of a stepped care intervention for alcohol use disorders in primary care: Pilot study. *British*

Journal of Psychiatry, 195, 448-456. doi:10.1192/bjp.bp.108.056697

Dunn, E. C., Wewiorski, N. J., & Rogers, E. S. (2008). The meaning and importance of employment to people in recovery from serious mental illness: Results of a qualitative study. *Psychiatric Rehabilitation Journal, 32,* 59-62. doi:10.2975/32.1.2008.59.62

Evans, C., Connell, J., Barkham, M., Margison, F., McGrath, G., Mellor-Clark, J., & Audin, K. (2002). Towards a standardised brief outcome measure: Psychometric properties and utility of the CORE-OM. *British Journal of Psychiatry, 180,* 51-60. doi:10.1192/bjp.180.1.51

Field, A. (2009). *Discovering statistics using SPSS* (3rd ed.). London, UK: Sage.

Firth, N., Barkham, M., Kellett, S., & Saxon, S. (2015). Therapist effects and moderators of effectiveness and efficiency in psychological wellbeing practitioners: A multilevel modelling analysis. *Behaviour Research and Therapy, 69,* 54-62. doi:10.1016/j.brat.2015.04.001

Frank, E., Rucci, P., Katon, W., Barrett, J., Williams, J. W., Oxman, T., … Cornell, J. (2002). Correlates of remission in primary care patients treated for minor depression. *General Hospital Psychiatry, 24,* 12-19. doi:10.1016/s0163-8343(01)00173-6

Gondek, D., Edbrooke-Childs, J., Fink, E., Deighton, J., & Wolpert, M. (2016). Feedback from outcome measures and treatment effectiveness, treatment efficiency, and collaborative practice: A systematic review. *Administration and Policy in Mental Health, 43,* 325-343. doi:10.1007/s10488-015-0710-5

Gratz, K. L., Dixon-Gordon, K. L., & Tull, M. T. (2014). Predictors of treatment response to an adjunctive emotion regulation group therapy for deliberate self-harm among women with borderline personality disorder. *Personality Disorders: Theory, Research, and Treatment, 5,* 97-107.

doi:10.1037/per0000062

Gratz, K. L., & Gunderson, J. G. (2006). Preliminary data on an acceptance-based emotion regulation group intervention for deliberate self-harm among women with borderline personality disorder. *Behavior Therapy, 37,* 25-35. doi:10.1016/j.beth.2005.03.002

Gratz, K. L., & Tull, M. T. (2011). Extending research on the utility of an adjunctive emotion regulation group therapy for deliberate self-harm among women with borderline personality pathology. *Personality Disorders: Theory, Research, and Treatment, 2,* 316-326. doi:10.1037/a0022144

Gratz, K. L., Tull, M. T., & Levy, R. (2013). Randomized controlled trial and uncontrolled 9-month follow-up of an adjunctive emotion regulation group therapy for deliberate self-harm among women with borderline personality disorder. *Psychological Medicine, 44,* 2099-2112. doi:10.1017/s0033291713002134

Green, H., Barkham M., Kellett, S., & Saxon, D. (2014). Therapist effects and IAPT psychological wellbeing practitioners (PWPs): A multilevel modelling and mixed methods analysis. *Behaviour Research and Therapy, 63,* 43-54. doi:10.1016/j.brat.2014.08.009

Guthrie, E., Kapur, N., Mackway-Jones, K., Chew-Graham, C., Moorey, J., Mendel, E., … Boddy, G. (2003). Predictors of outcome following brief psychodynamic-interpersonal therapy for deliberate self poisoning. *Australian and New Zealand Journal of Psychiatry, 37,* 532-536. doi:10.1046/j.1440-1614.2003.01197.x

Gyani, A., Shafran, R., Layard, R., & Clark, D. M. (2013). Enhancing recovery rates: Lessons from year one of IAPT. *Behaviour Research and Therapy, 51,* 597-606. doi:10.1016/j.brat.2013.06.004

Hawton, K., Townsend, E., Arensman, E., Gunnell, D., Hazell, P., House, A., &

van Heeringen, K. (1999). Psychosocial and pharmacological treatments for deliberate self-harm. *Cochrane Database of Systematic Reviews, 1999*(4), 1-42. doi:10.1002/14651858.cd001764.pub2

Hawton, K., Witt, K. G., Taylor Salisbury, T. L., Arensman, E., Gunnell, D., Hazell, P., … van Heeringen, K. (2016). Psychosocial interventions following self-harm in adults: A systematic review and meta-analysis. *Lancet Psychiatry, 3,* 740-750. doi:10.1016/s2215-0366(16)30070-0

Hetrick, S. E., Robinson, J., Spittal, M. J., & Carter, G. (2016). Effective psychological and psychosocial approaches to reduce repetition of self-harm: A systematic review, meta-analysis and meta-regression. *British Medical Journal Open, 6.* Retrieved from http://bmjopen.bmj.com/content/bmjopen/6/9/e011024.full.pdf

Hofmann, D. A., & Gavin, M. B. (1998). Centering decisions in hierarchical linear models: Implications for research in organizations. *Journal of Management, 24,* 623–641. doi:10.1177/014920639802400504

Hunt, I. M., Rahman, M. S., While, D., Windfuhr, K., Shaw, J., Appleby, L., & Kapur, N. (2014). Safety of patients under the care of crisis resolution home treatment services in England: A retrospective analysis of suicide trends from 2003 to 2011. *The Lancet Psychiatry, 1,* 135-141. doi:10.1016/S2215-0366(14)70250-0

Husain, N., Afsar, S., Ara, J., Fayyaz, H., Rahman, R. U., Tomenson, B., … Chaudhry, I. B. (2014). Brief psychological intervention after self-harm: Randomised controlled trial from Pakistan. *British Journal of Psychiatry, 204,* 462-470. doi:10.1192/bjp.bp.113.138370

IBM Corporation. (2015). *IBM SPSS Statistics for Windows, Version 23.0*. Armonk, NY: Author.

Jacobson, N. S., Follette, W. C., & Revenstorf, D. (1984). Psychotherapy outcome research: Methods for reporting variability and evaluating clinical

significance. *Behavior Therapy, 15,* 336-352. doi:10.1016/s0005-7894(84)80002-7

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

James, K., Stewart, D., & Bowers, L. (2012). Self-harm and attempted suicide within inpatient psychiatric services: A review of the literature. *International Journal of Mental Health Nursing, 21,* 301-309. doi:10.1111/j.1447-0349.2011.00794.x

Kim, D., Wampold, B. E., & Bolt, D. M. (2006). Therapist effects in psychotherapy: A random-effects modeling of the National Institute of Mental Health Treatment of Depression Collaborative Research Program data. *Psychotherapy Research, 16,* 161-172. doi:10.1080/10503300500264911

Knopp, J., Knowles, S., Bee, P., Lovell, K., & Bower, P. (2013). A systematic review of predictors and moderators of response to psychological therapies in OCD: Do we have enough empirical evidence to target treatment? *Clinical Psychology Review, 33,* 1067-1081. doi:10.1016/j.cpr.2013.08.008

Langner, J., Laws, M., Röper, G., Zaudig, M., Hauke, W., & Piesbergen, C. (2009). Predicting therapy outcome in patients with early and late obsessive-compulsive disorder (EOCD and LOCD). *Behavioural and Cognitive Psychotherapy, 37,* 485-496. doi:10.1017/S1352465809990294

Linehan, M. M. (1997). Behavioral treatments of suicidal behaviors. Definitional obfuscation and treatment outcomes. *Annals of the New York Academy of Sciences, 836,* 302-328. doi:10.1111/j.1749-6632.1997.tb52367

Lumley, T., Diehr, P., Emerson, S., & Chen, L. (2002). The importance of the normality assumption in large public health data sets. *Annual Review of*

Public Health, 23, 151-169. doi:10.1146/annurev.publhealth.23.100901.140546

Lutz, W., Leon, S. C., Martinovich, Z., Lyons, J. S., & Stiles, W. B. (2007). Therapist effects in outpatient psychotherapy: A three-level growth curve approach. *Journal of Counseling Psychology, 54,* 32-39. doi:10.1037/0022-0167.54.1.32

Lyne, K. J., Barrett, P., Evans, C., & Barkham, M. (2006). Dimensions of variation on the CORE-OM. *British Journal of Clinical Psychology, 45,* 185-203. doi:10.1348/014466505X39106

Maas, C. J. M., & Hox, J. J. (2005). Sufficient sample sizes for multilevel modelling. *Methodology, 1,* 86-92. doi:10.1027/1614-1881.1.3.86

Mann, J. J., Waternaux, C., Haas, G. L., & Malone, K. M. (1999). Toward a clinical model of suicidal behavior in psychiatric patients. *American Journal of Psychiatry, 156,* 181-189. Retrieved from http://ajp.psychiatryonline.org/doi/pdf/10.1176/ajp.156.2.181

McMillan, D., & Morley, S. (2010). Single case quantitative methods for practice-based evidence. In M. Barkham, G. E. Hardy & J. Mellor-Clark (Eds.), *Developing and delivering practice-based evidence: A guide for psychological therapies* (pp. 109-138). Chichester, UK: Wiley.

Mellor-Clark, J., & Barkham, M. (2000). Quality evaluation: Methods, measures and meaning. In C. Feltham & J. Horton (Eds.), *Handbook of counselling and psychotherapy* (pp. 225-270). London, UK: Sage.

Mellor-Clark, J., Barkham, M., Connell, J., & Evans, C. (1999). Practice-based evidence and standardized evaluation: Informing the design of the CORE system. *European Journal of Psychotherapy, Counselling and Health, 2,* 357-374. doi:10.1080/13642539908400818

Mental Health Taskforce. (2016). *The five year forward view for mental health.* London, UK: NHS England. Retrieved from https://www.england.nhs.uk/wp-content/uploads/2016/02/Mental-Health-Taskforce-FYFV-final.pdf

Milner, A., Page, A., & LaMontagne, A. D. (2014). Cause and effect in studies on unemployment, mental health and suicide: A meta-analytic and conceptual review. *Psychological Medicine, 44,* 909-917. doi:10.1017/S0033291713001621

Morris, S. B., & DeShon, R. P. (2002). Combining effect size estimates in meta-analysis with repeated measures and independent-groups designs. *Psychological Methods, 7,* 105-125. doi:10.1037//1082-989X.7.1.105

National Institute for Health and Care Excellence. (2011). *Commissioning stepped care for people with common mental health disorders. Commissioning guides CMG41.* Manchester, UK: Author. Retrieved from https://www.nice.org.uk/guidance/cmg41/chapter/3-a-stepped-care-approach-to-commissioning-high-quality-integrated-care-for-people-with-common

Nock, M. K. (2009). *Non-suicidal self-injury: Origins, assessment, and treatment*. Washington D. C., WA: American Psychological Association.

Norrie, J., Davidson, K., Tata, P., & Gumley, A. (2013). Influence of therapist competence and quantity of cognitive behavioural therapy on suicidal behaviour and inpatient hospitalisation in a randomised controlled trial in borderline personality disorder: Further analyses of treatment effects in the BOSCOT study. *Psychology and Psychotherapy: Theory, Research and Practice, 86,* 280-293. doi:10.1111/papt.12004

Okiishi, J. C., Lambert, M. J., Eggett, D., Nielson, S. L., Vermeersch, D. A., & Dayton, D. D. (2006). An analysis of therapist treatment effects: Toward

providing feedback to individual therapists on their patients' psychotherapy outcome. *Journal of Clinical Psychology, 62,* 1157–1172. doi:10.1002/jclp.20272

Okiishi, J. C., Lambert, M. J., Nielsen, S. L., & Ogles, B. M. (2003). Waiting for supershrink: An empirical analysis of therapist effects. *Clinical Psychology & Psychotherapy, 10,* 361-373. doi:10.1002/cpp.383

Oosterbaan, D. B., Verbraak, M. J., Terluin, B., Hoogendoorn, A. W., Peyrot, W. J., Muntingh, A., & van Balkom, A. J. (2013). Collaborative stepped care v. care as usual for common mental disorders: 8-month, cluster randomised controlled trial. *British Journal of Psychiatry, 203,* 132-139. doi:10.1192/bjp.bp.112.125211

Orlinsky, D. E., Rønnestad, M. H., & Willutzki, U. (2003). Fifty years of process-outcome research: Continuity and change. In M. J. Lambert (Ed.), *Handbook of psychotherapy and behavior change* (5th ed., pp. 307-390). New York, NY: Wiley.

Rasbash, J., Charlton, C., Browne, W. J., Healy, M., & Cameron, B. (2016). *MLwiN version 2.36.* Bristol, UK: Centre for Multilevel Modelling.

Rasbash, J., Steele, F., Browne, W. J., & Goldstein, H. (2016). *A user's guide to MLwiN version 2.36.* Bristol, UK: Centre for Multilevel Modelling.

Richards, D. A., & Suckling, R. (2009). Improving access to psychological therapies: Phase IV prospective cohort study. *British Journal of Clinical Psychology, 48,* 377-396. doi:10.1348/014466509X405178

Romano, M., & Peters, L. (2015). Evaluating the mechanisms of change in motivational interviewing in the treatment of mental health problems: A review and meta-analysis. *Clinical Psychology Review, 38,* 1-12. doi:10.1016/j.cpr.2015.02.008

Ryan, R. M., Lynch, M. F., Vansteenkiste, M., & Deci, E. L. (2011). Motivation and

autonomy in counseling, psychotherapy, and behavior change: A look at theory and practice. *The Counseling Psychologist, 39,* 193-260. doi:10.1177/0011000009359313

Saxon, D., & Barkham, M. (2012). Patterns of therapist variability: Therapist effects and the contribution of patient severity and risk. *Journal of Consulting and Clinical Psychology, 80,* 535-546. doi:10.1037/a0028898

Saxon, D., Firth, N., & Barkham, M. (2016). The relationship between therapist effects and therapy delivery factors: Therapy modality, dosage, and non-completion. *Administration and Policy in Mental Health and Mental Health Services Research.* Retrieved from https://link.springer.com/article/10.1007/s10488-016-0750-5/fulltext.html

Snijders, T. A. B., & Bosker, R. J. (1999). *Multilevel analysis.* Newbury Park, CA: Sage.

Soldz, S. (2006). Models and meanings: Therapist effects and the stories we tell. *Psychotherapy Research, 16,* 173–177. doi:10.1080/10503300500264937

Stiles, W. B., Barkham, M., & Wheeler, S. (2015). Duration of psychological therapy: Relation to recovery and improvement rates in UK routine practice. *British Journal of Psychiatry, 207,* 115-122. doi:10.1192/bjp.bp.114.145565

Tyron, G. S., Blackwell, S. C., & Hammel, E. F. (2007). A meta-analytic examination of client-therapist perspectives of the working alliance. *Psychotherapy Research, 17,* 629-642. doi:10.1080/10503300701320611

Vall, E., & Wade, T. D. (2015). Predictors of treatment outcome in individuals with eating disorders: A systematic review and meta-analysis. *International Journal of Eating Disorders, 48,* 946-971. doi:10.1002/eat.22411

Wampold, B. E., & Serlin, R. C. (2000). The consequences of ignoring a nested factor on measures of effect size in analysis of variance. *Psychological Methods, 5,*

425–433. doi:10.1037/1082-989X.5.4.425

Whewell, P., & Bonanno, D. (2000). The care programme approach and risk assessment of borderline personality disorder: Clinical validation of the CORE risk subscale. *Psychiatric Bulletin, 24,* 381-384. doi:10.1192/pb.24.10.381

Wolitzky-Taylor, K. B., Arch, J. J., Rosenfield, D., & Craske, M. G. (2012). Moderators and non-specific predictors of treatment outcome for anxiety disorders: A comparison of cognitive behavioral therapy to acceptance and commitment therapy. *Journal of Consulting and Clinical Psychology, 80,* 786-799. doi:10.1037/a0029418

Wunderlink, L., Nieboer, R. M., Wiersma, D., Sytema, S., & Nienhuis, F. J. (2013). Recovery in remitted first-episode psychosis at 7 years of follow-up of an early dose reduction/discontinuation or maintenance treatment strategy: Long-term follow-up of a 2-year randomized clinical trial. *Journal of the American Medical Association Psychiatry, 70,* 913-920. doi:10.1001/jamapsychiatry.2013.19

Appendix A

Down & Black's (1998) Quality Checklist

Removed for copyright purposes.

Appendix B

Ethical Approval Documents

NHS
Health Research Authority

NRES Committee Yorkshire & The Humber - Leeds East
Jarrow REC Centre
Room 002
Jarrow Business Centre
Rolling Mill Road
Jarrow
Tyne and Wear
NE32 3DT

Tel: 0191 428 3387

24 February 2014

Professor Michael Barkham
Professor of Clinical and Counselling Psychology
University of Sheffield
Centre for Psychological Services Research
Department of Psychology
Western Bank
Sheffield
S10 2TN

Dear Professor Barkham

Study title:	**An evaluation of the effectiveness of the psychological therapies as delivered in routine practice settings within primary and NHS service settings.**
REC reference:	**05/Q1206/128**
Amendment number:	**Amendment 3, 05/02/14**
Amendment date:	**05 February 2014**

The above amendment was reviewed by the Sub-Committee in correspondence.

Ethical opinion

It was noted that data had been gathered for some time, and clarification was requested as to whether the researchers had, or intended to, disseminate any interim results to participants.

You replied that because the dataset is anonymised, you cannot, de facto, disseminate interim results to participants. You had published extensively on the dataset such that the yield is in the scientific and public domain. Accordingly, you had disseminated findings in the way you envisaged but it would never have been possible to direct these to participants due to anonymity. You added that you could provide a listing of these publications if that would be of assistance to the Committee in arriving at a decision.

In addition to this, it was questioned whether the researchers were still anticipating using their previous plans for dissemination with this current data set, or if they had any new plans as the additional data set would be quite considerable.

You responded that because the dataset was anonymised at the level of the individual participant and the organisation (other than type of organisation), the same restrictions would apply in that you cannot feedback directly to participants at any level. However, the key difference in the additional data set was that the sectors from which the data was drawn, was broader. In that sense, it therefore enabled you to make comparisons between organisational sectors that were not possible in the earlier data set. Hence, while the data set is large in terms

A Research Ethics Committee established by the Health Research Authority

Figure B1. Page one (of two) of the ethical approval form.

of the total number of participants, when it was clustered according to the differing types of organisations (i.e., NHS [primary, secondary, tertiary], voluntary sector, university counselling, workplace counselling), then the number of participants within each type was considerably reduced. It was this diversity of organisational type that was the rationale for the current data set. In light of this, your dissemination plans would target learned journals and conference presentations focusing on organisational components.

You went on to say that an initial focus would be on the voluntary sector (VS) services as there is a paucity of research relating to this sector. Although the organisations in the data set are anonymous, you were planning on contacting local VS services to report on data from these analyses in order to provide them with the evidence arising from the data. Hence, there would be a direct means of dissemination to this sector. In other words, building a link between the yield of the data and 'real' organisations in the community, which might encourage them to collect and use data in support of their services. In this sense, the dissemination strategy was different as it could focus on organisational type in a way that the previous data sets did not.

The members of the Committee taking part in the review gave a favourable ethical opinion of the amendment on the basis described in the notice of amendment form and supporting documentation.

Approved documents

The documents reviewed and approved at the meeting were:

Document	Version	Date
Notice of Substantial Amendment (non-CTIMPs)	Amendment 3, 05/02/14	05 February 2014

Membership of the Committee

The members of the Committee who took part in the review are listed on the attached sheet.

R&D approval

All investigators and research collaborators in the NHS should notify the R&D office for the relevant NHS care organisation of this amendment and check whether it affects R&D approval of the research.

Statement of compliance

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

We are pleased to welcome researchers and R & D staff at our NRES committee members' training days – see details at http://www.hra.nhs.uk/hra-training/

05/Q1206/128:	**Please quote this number on all correspondence**

Yours sincerely

pp
Mrs Alison Barraclough
Chair

E-mail: nrescommittee.yorkandhumber-leedseast@nhs.net

Figure B2. Page two (of two) of the ethical approval form.

---------- Forwarded message ----------
From: **Amrit Sinha** <a.sinha@sheffield.ac.uk>
Date: 1 September 2016 at 11:18
Subject: Re: Fwd: ethics approval documentation
To: Katie Gregsoncurtis <kgregsoncurtis1@sheffield.ac.uk>

Hi Katie,

Thanks for sending this confirmation email through. This should suffice for evidence directly linking your study to the pre approved ethics. I will be able to issue your sponsor governance confirmation shortly.

Best wishes,

Amrit

On 29 August 2016 at 16:41, Katie Gregsoncurtis <kgregsoncurtis1@sheffield.ac.uk> wrote:

> Hi Amrit
>
> Please see email below from Michael and the attached document re: ethics.
>
> Please let me know if you need any other information.
>
> Many thanks
>
> Katie

---------- Forwarded message ----------
From: **Michael Barkham** <m.barkham@sheffield.ac.uk>
Date: 22 August 2016 at 08:49
Subject: Re: Fwd: ethics approval documentation
To: Katie Gregsoncurtis <kgregsoncurtis1@sheffield.ac.uk>

Katie

Good to hear from you.

Attached is the substantial amendment (favourable opinion) I received in 2014 for analyses on the 'new' CORE dataset, which is the one you are working on.

There is no governance as the data is annonymous so it is not possible to approach any R&D department and it doesn't need any as it is not utilizing any Trust resources.

Would be good to meet up sometime.

Keep me posted.

Best wishes

Michael

Figure B3. Email thread relating to ethical approval form.

Appendix C

Full Results of Analyses Investigating Differences on Demographic and Clinical Variables Between Study Samples and Excluded Patients

Table C1

Independent t-tests Comparing Data on Continuous Variables Between Excluded Patients and the Study-Specific Dataset

	Excluded patients			Study-specific dataset					
Variable	*n*	*M*	*SD*	*n*	*M*	*SD*	*t**	df	*d*
Age (years)	69,587	35.21	13.00	29,149	37.84	12.82	-29.30	55357.07	0.20
Number of problems	56,274	4.21	3.21	27,035	3.63	1.76	33.66	81905.77	0.00
Therapy attendance	35,314	0.76	0.26	27,470	0.90	0.16	-83.76	58991.24	0.65

Note. *d* = Cohen's *d*. Equal variances not assumed for any variable.

* all *p*s < .001

Table C2

Pearson's Chi-Square Analyses Comparing Data on Categorical Variables Between Excluded Patients and the Study-Specific Dataset

		Excluded patients		Study-specific dataset				
Variable	Category	Obs.	Exp.	Obs.	Exp.	χ^2*	df	*V*
Gender	Female	48,867	49,378.5	20,098	19,586.5	56.36	1	.02
	Male	24,939	24,427.5	9,178	9,689.5			
Employment	Employed/student	45,644	46,827.6	22,618	21,434.4	491.09	2	.08
	Unemployed	12,885	11,681.2	4,143	5,346.8			
	Retired	1,823	1,843.3	864	843.7			
Problem type(s)	Depression	34,215	35,208.8	15,578	14,584.2	7464.04	14	.15
	Anxiety/stress	40,675	43,230.9	20,463	17,907.1			
	Psychosis	4,040	2,971.2	162	1,230.8			
	Personality problems	8,019	6,320.1	919	2,617.9			
	Cognitive/learning	4,446	3,296.5	216	1,365.5			
	Eating disorder	5,414	4,516.3	973	1,870.7			
	Physical problems	10,276	10,076.2	3,974	4,173.8			
	Addiction	6,995	5,878.9	1,319	2,435.1			
	Trauma/abuse	14,001	13,258.9	4,750	5,492.1			
	Loss/bereavement	16,916	17,715.8	8,138	7,338.2			
	Self-esteem	25,031	26,635.9	12,628	11,033.1			
	Interpersonal	31,547	32,614.4	14,577	13,509.6			
	Welfare/living	11,092	10,632.0	3,944	4,404.0			
	Work/academic	17,029	18,063.0	8,516	7,482.0			

(continued)

Variable	Category	Excluded patients		Study specific dataset				
		Obs.	Exp.	Obs.	Exp.	χ^2*	df	*V*
Ethnicity	Asian	4,029	3,580.2	1,172	1,620.8	1124.30	4	.11
	Mixed race	614	560.3	200	253.7			
	Black	3,447	2,993.7	902	1,355.3			
	White	53,948	54,024.3	24,533	24,456.7			
	Other/not stated	2,632	3,511.4	2,469	1,589.6			
Therapy modality	Integrative	17,342	18,845.1	16,505	15,001.9	595.54	1	.10
	Other	17,184	15,680.9	10,980	12,483.1			
Therapy ending	Planned	17,512	24,211.3	25,475	18,775.7	13119.51	1	.45
	Unplanned	18,277	11,577.7	2,279	8,978.3			
Service type	Voluntary	15,190	15,696.1	6,617	6,110.9	2578.26	5	.16
	University/workplace	30,974	31,238.1	12,426	12,161.9			
	Primary care	18,341	19,443.2	8,672	7,569.8			
	Secondary care	6,189	5,288.9	1,159	2,059.1			
	Tertiary care	4,354	3,216.7	115	1,252.3			
	Private	151	316.0	288	123.0			

Note. Obs. = Observed. Exp. = Expected. *V* = Cramer's *V*. Patients could report multiple problem types.

* all *p*s < .001

Table C3

Independent t-tests Comparing Data on Continuous Variables Between Study-Specific Patients Included and Not Included in the Multilevel Modelling Analysis

Variable	Non-MLM			MLM					
	n	*M*	*SD*	*n*	*M*	*SD*	*t*	df	*d*
Age (years)	24,176	37.72	12.78	4,974	38.45	13.07	-3.63*	7065.26	-0.06
Number of problems	24,302	3.23	1.96	4,975	3.95	1.80	-25.28*	7579.88	-0.38
Therapy attendance	22,535	0.90	0.16	4,935	0.90	0.15	-69.00	7586.16	-0.01

Note. d = Cohen's *d*. Equal variances not assumed for any variable. MLM = Sample with which the final multilevel model was developed.

* *p* < .001

Table C4

Pearson's Chi-Square Analyses Comparing Data on Categorical Variables Between Study-Specific Patients Included and Not Included in the Multilevel Modelling Analysis

Variable	Category	Non-MLM		MLM				
		Obs.	Exp.	Obs.	Exp.	χ^2	df	*V*
Gender	Female	16,704	16,682.9	3,395	3,416.1	0.50	1	.00
	Male	7,597	7,618.1	1,581	1,559.9			
Employment	Employed/student	18,715	18,544.9	3,904	4,074.1	47.90*	2	.04
	Unemployed	3,253	3,396.8	890	746.2			
	Retired	682	708.4	182	155.6			

(continued)

Variable	Category	Non-MLM		MLM		χ^2	df	*V*
		Obs.	Exp.	Obs.	Exp.			
Problem type(s)	Depression	12,255	12,458.8	3,323	3,119.2	179.98*	14	.04
	Anxiety/stress	16,647	16,367.2	3,818	4,097.8			
	Psychosis	129	123.2	25	30.8			
	Personality problems	797	731.0	117	183.0			
	Cognitive/learning	175	167.2	34	41.8			
	Eating disorder	760	773.4	207	193.6			
	Physical problems	3,077	3,167.9	884	793.1			
	Addiction	1,022	1,055.7	298	264.3			
	Trauma/abuse	3,708	3,798.1	1,041	950.9			
	Loss/bereavement	6,402	6,510.9	1,739	1,630.1			
	Self-esteem	10,079	10,100.3	2,550	2,528.7			
	Interpersonal	11,807	11,659.0	2,771	2,919.0			
	Welfare/living	3,062	3,154.3	882	789.7			
	Work/academic	6,842	6,810.0	1,673	1,705.0			
	Other	1,636	1,521.2	266	380.8			
Ethnicity	Asian	919	972.8	253	199.2	205.50*	4	.08
	Mixed race	164	166.0	36	34.0			
	Black	741	748.7	161	153.3			
	White	20,179	20,364.1	4,355	4,169.9			
	Other/not stated	2,298	2,049.4	171	419.6			
Therapy modality	Integrative	13,072	13,520.8	3,434	2,985.2	206.19*	1	.09
	Other	9,443	8,994.2	1,537	1,985.8			
Therapy ending	Planned	20,865	20,916.8	4,612	4,560.2	8.72**	1	.02
	Unplanned	1,922	1,870.2	356	407.8			
Service type	Voluntary	6,356	5,492.4	261	1,124.6	1860.93*	5	.25
	University/workplace	10,344	10,314.9	2,083	2,112.1			
	Primary care	6,150	7,198.1	2,522	1,473.9			
	Secondary care	1,111	962.0	48	197.0			
	Tertiary care	115	95.5	0	19.5			
	Private	226	239.1	62	48.9			

Note. MLM = Sample with which the final multilevel model was developed. Obs. = Observed. Exp. = Expected. *V* = Cramer's *V*. Patients could report multiple problem types.

* $p < .001$ ** p = .003

Appendix D

Clinical Outcomes in Routine Evaluation-Outcome Measure

Removed for copyright purposes.

Appendix E

Clinical Outcomes in Routine Evaluation-Therapy Assessment Form

Removed for copyright purposes.

Appendix F

Clinical Outcomes in Routine Evaluation-End of Therapy Form

Removed for copyright purposes.

Appendix G

Nonparametric Analyses

Nonparametric Analyses of Variability in Risk-to-Self and Non-Risk Scores Between Services

Pre-therapy Clinical Outcomes in Routine Evaluation-Outcome Measure (CORE-OM) scores were compared between patients accessing different service types. A Kruskal-Wallis test indicated that there was a significant difference in CORE-OM risk-to-self scores across services, $H(5) = 544.98$, $p < .001$. Significant differences in risk-to-self scores between each service type were investigated using Mann Whitney *U* tests (see Table G1). Results indicated that secondary care patients had significantly higher risk-to-self scores than patients in any other service type. Patients in university/workplace services reported significantly lower risk-to-self scores than patients in voluntary, primary, and tertiary care services.

Table G1

Mann Whitney U Comparisons of Risk-to-Self Scores Across Services

Compared services	Mean ranks	*U*	*p*
Voluntary vs. University/workplace	10025.79 vs. 9253.72	37777826.00	<.001[a]
Voluntary vs. Primary	7652.16 vs. 7639.54	28643941.50	.426[a]
Voluntary vs. Secondary	3725.32 vs. 4820.15	2754767.00	<.001[a]
Voluntary vs. Tertiary	3361.61 vs. 3647.78	348130.00	.047[a]
Voluntary vs. Private	3467.07 vs. 3129.65	859722.00	.003[b]
University/workplace vs. Primary	10206.56 vs. 11040.89	49617825.50	<.001[a]
University/workplace vs. Secondary	6589.41 vs. 8975.72	4671092.00	<.001[a]
University/workplace vs. Tertiary	6261.63 vs. 7283.34	598075.50	<.001[a]
University/workplace vs. Private	6360.01 vs. 6249.29	1758180.00	.574[b]
Primary vs. Secondary	4752.99 vs. 6135.68	3611820.50	<.001[a]
Primary vs. Tertiary	4389.06 vs. 4766.75	455774.00	.045[a]
Primary vs. Private	4494.25 vs. 4066.44	1129517.50	.003[b]
Secondary vs. Tertiary	649.38 vs. 517.75	52871.00	<.001[a]
Secondary vs. Private	776.54 vs. 512.56	106001.50	<.001[b]
Tertiary vs. Private	227.80 vs. 191.70	13593.00	.002[b]

[a] one-tailed test. [b] two-tailed test.

A Kruskal-Wallis test indicated that there was a significant difference in CORE-OM non-risk scores across services, $H(5) = 332.27$, $p < .001$. Significant differences in

non-risk scores between each service type were investigated using Mann Whitney *U* tests (see Table G2). Results indicated that secondary care patients had significantly higher non-risk scores than patients in university/workplace, voluntary, primary care, and private services, and marginally significantly higher non-risk scores than tertiary care patients. Patients in voluntary services had significantly lower non-risk scores than patients in university/workplace or primary care services. Patients in university/workplace services had significantly lower non-risk scores than patients in primary care services.

Table G2

Mann Whitney U Comparisons of Non-Risk Scores Across Services

Compared services	Mean ranks	*U*	*p* (two-tailed)
Voluntary vs. University/workplace	9317.75 vs. 9630.76	39759921.50	<.001
Voluntary vs. Primary	7115.00 vs. 8049.41	25184301.50	<.001
Voluntary vs. Secondary	3749.24 vs. 4683.56	2913075.50	<.001
Voluntary vs. Tertiary	3363.40 vs. 3544.70	359985.00	.321
Voluntary vs. Private	3445.71 vs. 3620.47	904615.50	.145
University/workplace vs. Primary	10142.78 vs. 11132.28	48825270.50	<.001
University/workplace vs. Secondary	6666.61 vs. 8148.07	5630340.50	<.001
University/workplace vs. Tertiary	6269.84 vs. 6395.89	700133.00	.710
University/workplace vs. Private	6345.59 vs. 6482.89	1753233.00	.558
Primary vs. Secondary	4839.33 vs. 5489.66	4360551.00	<.001
Primary vs. Tertiary	4398.19 vs. 4077.67	462262.00	.178
Primary vs. Private	4490.83 vs. 4169.36	1159158.50	.038
Secondary vs. Tertiary	648.72 vs. 524.41	53637.00	.001
Secondary vs. Private	752.50 vs. 609.29	133859.50	<.001
Tertiary vs. Private	201.96 vs. 202.02	16555.00	.996

Nonparametric Comparisons of Scores on Continuous Predictor Variables Between Patients Who Did and Did Not Meet Criteria for RCSI

Differences between the scores of patients who did and did not meet criteria for RCSI on potential continuous predictor variables were analysed using Mann Whitney *U* tests (see Table G3). The results indicated that patients who met criteria for RCSI had significantly lower CORE-OM risk and non-risk scores than patients who did not meet criteria for RCSI. These differences had large effect sizes. Effect sizes for differences on other variables were very small.

Table G3

Mann Whitney U Comparisons of Scores on Continuous Predictor Variables Between Patients Who Did and Did Not Meet Criteria for RCSI

Variable	Mean rank		*U*	*p*	η^2
	RCSI group	No RCSI group			
Age	6701.82	6974.67	22355831.50	<.001	.001
Number of presenting problems	6078.52	6563.19	18312103.50	<.001	.005
Attendance	6639.11	6210.12	19273022.50	<.001	.004
Referral to assessment wait	5785.42	6344.52	16547139.50	<.001	.006
Therapy length	6337.92	6423.68	20008237.50	.19	.000
Episode of mental health difficulty	6810.91	6911.48	23130433.00	<.001	.001
Pre-therapy CORE-OM risk score	6310.89	7457.16	19551313.00	<.001	.021
Pre-therapy CORE-OM non-risk score	6094.09	7693.77	17999400.00	<.001	.041

Note. RCSI = reliable and clinically significant improvement. η^2 = eta-squared. CORE-OM = Clinical Outcomes in Routine Evaluation-Outcome Measure.

Appendix H

Pearson's Chi-Square Analyses with Original Categorical Variables

Variable	Category	No RCSI group		RCSI group		χ^2	df	*V*
		Obs.	Exp.	Obs.	Exp.			
Medication use in therapy	None throughout	2724	2855.9	3358	3226.1	238.64[a]	6	.14[a]
	Discontinued	124	216.0	336	244.0			
	Decreased	125	197.7	296	223.3			
	Modified	114	98.6	96	111.4			
	Increased	129	84.5	51	95.5			
	Started	200	174.7	172	197.3			
	Maintained	1993	1781.6	1801	2012.4			
Employment	Employed/student	4448	4764.0	5571	5255.0	194.21[a]	2	.12[a]
	Unemployed	1524	1211.1	1023	1335.9			
	Retired	170	166.9	181	184.1			
Carer role	None	60	47.9	46	58.1	5.91	2	.04
	Carer of children	1673	1686.6	2058	2044.4			
	Other caring role	30	28.5	33	34.5			
Therapy frequency	Less than weekly	1416	1557.0	1877	1736.0	55.07[a]	3	.07[a]
	Weekly	3824	3627.9	3849	4045.1			
	More than weekly	40	35.0	34	39.0			
	No fixed frequency	669	729.1	873	812.9			
Concurrent care	None	1684	1631.6	1654	1706.4	52.65[b]	4	.08[a]
	Voluntary care	34	33.7	35	35.3			
	Primary care	2435	2556.0	2794	2673.0			
	Secondary care	240	213.1	196	222.9			
	Specialist	148	106.6	70	111.4			
Previous care	None	2165	2305.5	2562	2421.5	70.90[a]	4	.09[a]
	Voluntary care	591	580.4	599	609.6			
	Primary care	833	846.7	903	889.3			
	Secondary care	404	342.9	299	360.1			
	Specialist	495	412.6	351	433.4			
Ethnicity	Asian	387	310.8	266	342.2	95.38[a]	4	.09[a]
	Mixed race	62	53.3	50	58.7			
	Black	250	208.4	188	229.6			
	White	5176	5371.9	6112	5916.1			
	Other	403	333.6	298	367.4			
Living status	Alone	1876	1683.3	1663	1855.7	74.46[a]	4	.08[a]
	With friends/family	3181	3384.7	3935	3731.3			
	Institution	68	53.3	44	58.7			
	Shared housing	795	802.4	892	884.6			
	Temporary housing	27	23.3	22	25.7			

Note. RCSI = reliable and clinically significant improvement. Obs. = Observed. Exp. = Expected. *V* = Cramer's *V*.

[a] Monte Carlo $p < .001$, 99% confidence intervals [.000, .000]. [b] Exact $p < .001$.

Appendix I

Full Results of Analyses to Determine Differences in Scores on Predictor Variables Between Patients Who Did and Did Not Meet Criteria for RCSI

Table I1

Independent t-tests Comparing Data on Continuous Predictor Variables Between Patients Who Did and Did Not Meet Criteria for RCSI

Predictor	No RCSI		RCSI		*t*	df	*d*
	M	*SD*	*M*	*SD*			
Age[a]	37.22	12.89	36.34	12.55	4.07*	13472.15	0.07
Number of presenting problems[a]	4.09	1.90	3.82	1.77	8.09*	12239.11	0.15
Attendance[a]	0.89	0.17	0.91	0.15	-7.27*	12273.94	0.13
Referral to assessment wait[a]	40.50	56.35	32.68	50.86	7.98*	11556.58	0.15
Therapy length	122.80	133.59	120.87	134.97	0.81	12754.00	0.01
Episode of mental health difficulty[a]	1.09	0.37	1.07	0.34	2.92	13323.92	0.06
Pre-therapy CORE-OM risk score[a]	8.59	6.62	6.71	5.50	17.95*	12788.71	0.31
Pre-therapy CORE-OM non-risk score	25.59	5.42	23.42	5.33	23.65*	13715.00	0.40

Note. *d* = Cohen's *d*. CORE-OM = Clinical Outcomes in Routine Evaluation-Outcome Measure. RCSI = reliable and clinically significant improvement.

[a] equal variances not assumed.

* $p < .001$.

Table I2

Pearson's Chi-Square Analyses of Scores on Categorical Predictor Variables for Patients Who Did and Did Not Meet Criteria for RCSI

Variable	Category	No RCSI		RCSI		χ^2	df	*V*
		Obs.	Exp.	Obs.	Exp.			
Gender	Male	2357	2286.1	2424	2494.9	6.47	1	.02
	Female	4202	4272.9	4734	4663.1			
Medication use in therapy	None throughout	2724	2855.9	3358	3226.1	220.64[a]	3	.14[a]
	Reduced	249	413.7	632	467.3			
	Increased/modified	443	357.8	319	404.2			
	Maintained	1993	1781.6	1801	2012.4			
Therapy type	Integrative	3540	3695.4	4264	4108.6	31.49[b]	1	.05[a]
	Other	2558	2402.6	2516	2671.4			
Employment	Employed/student	4448	4764.0	5571	5255.0	178.13[b]	1	.12[a]
	Unemployed	1694	1378.0	1204	1520.0			
Motivation	Poor	402	248.6	124	277.4	591.97[a]	2	.23[a]
	Moderate	1742	1327.4	1067	1481.6			
	Good	3098	3666.0	4660	4092.0			

(continued)

Variable	Category	No RCSI		RCSI				
		Obs.	Exp.	Obs.	Exp.	χ^2	df	*V*
Working alliance	Poor	247	150.2	71	167.8	498.84[a]	2	.21[a]
	Moderate	1657	1247.8	984	1393.2			
	Good	3333	3838.9	4792	4286.1			
Psychological mindedness	Poor	655	455.9	311	510.1	477.65[a]	2	.21[a]
	Moderate	2199	1854.6	1731	2075.4			
	Good	2372	2915.5	3806	3262.5			
Concurrent care	No	1684	1580.6	1654	1757.4	19.71[b]	1	.05[a]
	Yes	2857	2960.4	3395	3291.6			
Previous care	No	2165	2305.5	2562	2421.5	34.35[b]	1	.06[a]
	Yes	2323	2182.5	2152	2292.5			
Ethnicity	Other	1102	906.1	802	997.9	94.44[b]	1	.09[a]
	White	5176	5371.9	6112	5916.1			
Therapy frequency	Other	2125	2321.1	2784	2587.9	51.52[b]	1	.06[a]
	Weekly	3824	3627.9	3849	4045.1			
Living status	Alone	1876	1683.3	1663	1855.7	65.71[a]	2	.07[a]
	Friends/family	3181	3384.7	3935	3731.3			
	Other	890	879.0	958	969.0			
Caring role	No	60	47.9	46	58.1	5.72	1	.04
	Yes	1703	1715.1	2091	2078.9			
Problem duration	< 6 months	570	712.8	942	799.2	114.37[a]	3	.10[a]
	6-12 months	483	568.5	723	637.5			
	> 12 months	1262	1269.1	1430	1422.9			
	Recurring/continuous	3164	2928.5	3048	3283.5			

Note. Obs. = Observed. Exp. = Expected. *V* = Cramer's *V*. RCSI = reliable and clinically significant improvement.

[a] Monte Carlo significance $p < .001$, 99% confidence intervals [.000, .000]. [b] Exact significance $p < .001$.

Appendix J

Multicollinearity Analyses

Several analyses were undertaken to identify issues of multicollinearity between potential predictor variables. First, dummy variables were created for each categorical variable and each predictor variable was entered into a simple linear regression analysis to obtain tolerance and variance inflation factor statistics (see Field, 2009). The results of this analysis are presented in Table J1. Generally, tolerance values < 0.1 and variance inflation factor values > 10.00 suggest multicollinearity.

Table J1

Tolerance and Variance Inflation Factor Statistics for Potential Predictor Variables

Variable	Tolerance	VIF
Pre-therapy CORE-OM non-risk score	0.78	1.29
Pre-therapy CORE-OM risk score	0.70	1.44
Reduced/discontinued medication[a]	0.93	1.08
Increased/modified medication[a]	0.93	1.08
Maintained medication[a]	0.87	1.16
Moderate motivation[b]	0.16	6.30
Good motivation[b]	0.12	8.09
Employed/student[c]	0.93	1.07
Problem duration 6-12 months[d]	0.62	1.61
Problem duration > 12 months[d]	0.46	2.18
Problem duration - recurrent/continuous[d]	0.41	2.47
Moderate working alliance[e]	0.10	9.81
Good working alliance[e]	0.09	11.77
Moderate psychological mindedness[f]	0.23	4.28
Good psychological mindedness[f]	0.18	5.50

Note. VIF = variance inflation factor. CORE-OM = Clinical Outcomes in Routine Evaluation-Outcome Measure.
[a] reference category *no medication*. [b] reference category *poor* motivation. [c] reference category *unemployed.* [d] reference category < *6 months* problem duration. [e] reference category *poor* working alliance. [f] reference category *poor* psychological mindedness.

As Table J1 shows, the only variable which appeared at risk of multicollinearity was working alliance ratings. Predictors with high loadings on small eigenvalues were reviewed. Ratings of moderate and good working alliances both loaded highly (0.72 and 0.84, respectively) on the eigenvalue of .015. This suggested dependency between the ratings of working alliance.

To explore associations between continuous predictors, Pearson's correlation analyses were conducted. To explore associations between continuous and dichotomous predictors, point-biserial correlation analyses were conducted. A medium-sized, positive significant correlation was found between pre-therapy Clinical Outcomes in Routine Evaluation-Outcome Measure (CORE-OM) non-risk and pre-therapy CORE-OM risk scores, $r = .44$, $p < .001$. Small-sized positive correlations were found between employment status and pre-therapy CORE-OM risk scores, $r_{pb} = .12$, $p < .001$, and pre-therapy CORE-OM non-risk scores, $r_{pb} = .15$, $p < .001$.

The relationships between continuous variables and categorical variables with ≥ 3 categories were explored using analyses of variance (ANOVAs) and eta-squared effect sizes (see Table J2). The results indicated that none of the relationships between the continuous and categorical variables appeared overly large.

Table J2

Results of ANOVAs Between Continuous and Categorical Variables

Variables	*F*	*p*	η^2
Pre-therapy CORE-OM non-risk score vs Problem duration	8.07	<.001	0.003
Pre-therapy CORE-OM non-risk score vs Motivation	12.88	<.001	0.003
Pre-therapy CORE-OM non-risk score vs Working alliance	2.32	.099	0.003
Pre-therapy CORE-OM non-risk score vs Psychological mindedness	15.38	<.001	0.003
Pre-therapy CORE-OM non-risk score vs Medication	122.03	<.001	0.040
Pre-therapy CORE-OM risk score vs Problem duration	33.02	<.001	0.011
Pre-therapy CORE-OM risk score vs Motivation	37.14	<.001	0.008
Pre-therapy CORE-OM risk score vs Working alliance	12.77	<.001	0.000
Pre-therapy CORE-OM risk score vs Psychological mindedness	46.62	<.001	0.010
Pre-therapy CORE-OM risk score vs Medication	43.59	<.001	0.015

Note. η^2 = eta-squared

The relationships between each pair of categorical variables were explored using Pearson's chi-square analyses and Cramer's *V* effect sizes. The results of these analyses are presented in Table J3. Large sized positive relationships were found between ratings of working alliance, psychological mindedness, and motivation (all *p*s <.001). This result suggested potential issues with multicollinearity between therapist ratings of patients' psychological mindedness, motivation, and working alliances.

Table J3

Pearson's Chi-Square Analyses Between Pairs of Categorical Predictor Variables

Compared variables	χ^2	df	*p*	*V*
Medication vs Employment	345.59	1	<.001	.18
Medication vs Problem Duration	49.75	2	<.001	.11
Medication vs Psychological Mindedness	40.56	2	<.001	.05
Medication vs Working Alliance	14.27	2	.03	.05
Medication vs Motivation	60.08	2	<.001	.05
Employment vs Problem Duration	254.37	3	<.001	.15
Employment vs Psychological Mindedness	123.04	2	<.001	.11
Employment vs Working Alliance	37.40	2	<.001	.06
Employment vs Motivation	63.40	2	<.001	.08
Problem Duration vs Psychological Mindedness	83.60	2	<.001	.07
Problem Duration vs Working Alliance	53.04	2	<.001	.05
Problem Duration vs Motivation	59.89	2	<.001	.06
Psychological Mindedness vs Working Alliance	4859.63	2	<.001	.50
Psychological Mindedness vs Motivation	6986.37	2	<.001	.56
Working Alliance vs Motivation	4870.55	2	<.001	.47

Note. V = Cramer's *V*.

Appendix K

Full Single-Level Model

$$RCSI_i \sim Binomial(c18_i, \pi_i)$$

$$\begin{aligned} logit(\pi_i) = {} & 0.770(0.087)cons + -0.076(0.006)(PreDistress\text{-}gm)_i + \\ & -0.960(0.070)Poor/Moderate_i + 0.932(0.123)Reduced_i + \\ & -0.095(0.064)Inc/Chang/Maint_i + -0.290(0.096)>6mnths_i + \\ & -0.521(0.092)Recurrent/continuous_i \end{aligned}$$

$$var(RCSI_i|\pi_i) = \pi_i(1 - \pi_i)/c18_i$$

(4976 of 4976 cases in use)

Figure K1. Full single-level model of patient-level predictors of the likelihood of patients meeting criteria for reliable and clinically significant improvement (RCSI). c18 = denominator. cons = constant. PreDistress-gm = pre-therapy Clinical Outcomes in Routine Evaluation-Outcome Measure nonrisk score grand-mean centred. Poor/moderate = poor/moderate motivation. Reduced = reduced medication. Inc/Chang/Maint = increased/modified/maintained medication. >6mnths = > 6 months problem duration. Recurrent/continuous = recurrent/continuous problem duration.

Appendix L

Full Two-Level Model

$$RCSI_{ij} \sim \text{Binomial}(c18_{ij}, \pi_{ij})$$

$$\text{logit}(\pi_{ij}) = \beta_{0j}\text{cons} + -0.072(0.006)(\text{PreDistress-gm})_{ij} + -1.036(0.075)\text{Poor/Moderate}_{ij} + 0.750(0.126)\text{Reduced}_{ij} + -0.110(0.066)\text{Inc/Chang/Maint}_{ij} + -0.419(0.084)\text{Unemployed/Retired}_{ij}$$

$$\beta_{0j} = 0.548(0.087) + u_{0j}$$

$$\begin{bmatrix} u_{0j} \end{bmatrix} \sim N(0, \Omega_u) : \Omega_u = \begin{bmatrix} 0.418(0.079) \end{bmatrix}$$

$$\text{var}(RCSI_{ij}|\pi_{ij}) = \pi_{ij}(1 - \pi_{ij})/c18_{ij}$$

(4976 of 4976 cases in use)

Figure L1. Full two-level model of patient-level predictors and therapist effects in predicting the likelihood of patients meeting criteria for reliable and clinically significant improvement (RCSI). This model was developed using marginal quasi-likelihood approximation procedures. c18 = denominator. cons = constant. PreDistress-gm = pre-therapy Clinical Outcomes in Routine Evaluation-Outcome Measure nonrisk score grand-mean centred. Poor/moderate = poor/moderate motivation. Reduced = reduced medication. Inc/Chang/Maint = increased/modified/maintained medication.

Appendix M

Full Two-Level Model with Outlying Data Excluded

$$RCSI_{ij} \sim \text{Binomial}(c18_{ij}, \pi_{ij})$$

$$\text{logit}(\pi_{ij}) = \beta_{0j}\text{cons} + \underset{(0.006)}{-0.073}(\text{PreDistress-gm})_{ij} + \underset{(0.075)}{-1.040}\text{Poor/Moderate}_{ij} + \underset{(0.126)}{0.765}\text{Reduced}_{ij} + \underset{(0.067)}{-0.113}\text{Inc/Chang/Maint}_{ij} + \underset{(0.084)}{-0.437}\text{Unemployed/Retired}_{ij}$$

$$\beta_{0j} = \underset{(0.085)}{0.529} + u_{0j}$$

$$\begin{bmatrix} u_{0j} \end{bmatrix} \sim N(0, \Omega_u) \; : \; \Omega_u = \begin{bmatrix} \underset{(0.072)}{0.376} \end{bmatrix}$$

$$\text{var}(RCSI_{ij}|\pi_{ij}) = \pi_{ij}(1 - \pi_{ij})/c18_{ij}$$

(4919 of 4976 cases in use)

Figure M1. Full two-level model of patient-level predictors and therapist effects in predicting the likelihood of patients meeting criteria for reliable and clinically significant improvement (RCSI). This model did not include data from one outlier, who provided data for 57 patients. This model was developed using marginal quasi-likelihood approximation procedures. c18 = denominator. cons = constant. PreDistress-gm = pre-therapy Clinical Outcomes in Routine Evaluation-Outcome Measure nonrisk score grand-mean centred. Poor/moderate = poor/moderate motivation. Reduced = reduced medication. Inc/Chang/Maint = increased/modified/maintained medication.

Appendix N

Full Sensitivity Analysis Model

$$RCSI_{ij} \sim \text{Binomial}(\text{denominator}_{ij}, \pi_{ij})$$

$$\text{logit}(\pi_{ij}) = \beta_{0j}\text{cons} + \text{-0.073(0.005)PreDistress}_{ij} + \text{-1.473(0.126)Poor}_{ij} + \text{-0.920(0.056)Moderate}_{ij} + \text{0.802(0.096)Reduced}_{ij} + \text{-0.141(0.050)Increase/Maintain/Modify}_{ij} + \text{-0.440(0.060)Unemployed/Retired}_{ij}$$

$$\beta_{0j} = 2.319(0.115) + u_{0j}$$

$$\begin{bmatrix} u_{0j} \end{bmatrix} \sim N(0, \Omega_u) : \Omega_u = \begin{bmatrix} 0.338(0.046) \end{bmatrix}$$

$$\text{var}(RCSI_{ij}|\pi_{ij}) = \pi_{ij}(1 - \pi_{ij})/\text{denominator}_{ij}$$

(8885 of 8885 cases in use)

Figure N1. Full two-level model of patient-level predictors and therapist effects in predicting the likelihood of patients meeting criteria for reliable and clinically significant improvement (RCSI) using the whole therapist sample. This model was developed using marginal quasi-likelihood approximation procedures. cons = constant. PreDistress-gm = pre-therapy Clinical Outcomes in Routine Evaluation-Outcome Measure nonrisk score grand-mean centred. Poor = poor motivation. Moderate = moderate motivation. Reduced = reduced medication. Inc/Chang/Maint = increased/modified/maintained medication.