Accepted Manuscript

A multi-service practice research network study of large group psychoeducational cognitive behavioural therapy

Jaime Delgadillo, Stephen Kellett, Shehzad Ali, Dean McMillan, Michael Barkham, David Saxon, Gill Donohoe, Heather Stonebank, Sarah Mullaney, Patricia Eschoe, Richard Thwaites, Mike Lucock

PII: S0005-7967(16)30167-X

DOI: 10.1016/j.brat.2016.09.010

Reference: BRT 3038

To appear in: *Behaviour Research and Therapy*

Received Date: 5 March 2016

Revised Date: 23 June 2016

Accepted Date: 19 September 2016

Please cite this article as: Delgadillo, J., Kellett, S., Ali, S., McMillan, D., Barkham, M., Saxon, D., Donohoe, G., Stonebank, H., Mullaney, S., Eschoe, P., Thwaites, R., Lucock, M., A multi-service practice research network study of large group psychoeducational cognitive behavioural therapy, *Behaviour Research and Therapy* (2016), doi: 10.1016/j.brat.2016.09.010.

A multi-service practice research network study of large group psychoeducational cognitive behavioural therapy

Jaime Delgadillo [a*], Stephen Kellett [b,e], Shehzad Ali [c], Dean McMillan [d], Michael Barkham [b], David Saxon [b], Gill Donohoe [e], Heather Stonebank [e], Sarah Mullaney [f], Patricia Eschoe [f], Richard Thwaites [g], and Mike Lucock [h,f]

a. Leeds Community Healthcare NHS Trust and Department of Health Sciences, University of York, UK
b. Centre for Psychological Services Research, University of Sheffield, UK
c. Department of Health Sciences and Centre for Health Economics, University of York, UK
d. Department of Health Sciences and Hull York Medical School, University of York, UK
e. Sheffield Health & Social Care NHS Foundation Trust, UK
f. South West Yorkshire Partnership NHS Foundation Trust, UK
g. Cumbria Partnership NHS Foundation Trust, UK
h. Centre for Applied Psychological and Health Research, University of Huddersfield, UK

Word count (excl. tables and references): 4,627

Declarations of interest: None.

* Corresponding author.

E-mail address: jaime.delgadillo@nhs.net (J. Delgadillo).

Abstract

[word count: 176]

Background: This was a multi-service evaluation of the clinical and organisational effectiveness of large group psychoeducational CBT delivered within a stepped care model.

Method: Clinical outcomes for 4,451 participants in 163 psychoeducational groups delivered across 5 services were analysed by calculating pre-post treatment anxiety (GAD-7) effect sizes (Cohen's *d*). Overall and between-service effects were compared to published efficacy benchmarks. Multilevel modelling was used to examine if variability in clinical outcomes was explained by differences in service, group and patient-level (case-mix) variables.

Results: The pooled GAD-7 (pre-post) effect size for all services was $d = 0.70$, which was consistent with efficacy benchmarks for guided self-help interventions ($d = 0.69$). One service had significantly smaller effects ($d = 0.48$), which was explained by differences in group treatment length and case-mix. Variability between groups (i.e., *group effects*) explained up to 3.6% of variance in treatment outcomes.

Conclusions: Large group psychoeducational CBT is clinically effective, organisationally efficient and consistent with a stepped care approach to service design. Clinical outcome differences between services were largely explained by group and patient variables.

Key words: low intensity cognitive behavioural therapy; psychoeducation; depression; anxiety; IAPT; multilevel modelling

1. Introduction

In the United Kingdom, low intensity guided self-help cognitive behavioural interventions are a key feature of services within the *Improving Access to Psychological Therapies* (IAPT) programme (Clark, 2011). A commonly available low intensity treatment is the *Stress Control* (SC) programme (White & Keenan, 1990) delivered as an entry-level support option within stepped care IAPT service models. SC is a group-based didactic intervention that teaches anxiety and depression coping skills; it is delivered as a series of 6 lecture-style sessions based on principles of cognitive behavioural therapy (CBT). The content of SC is similar to other CBT-based self-help interventions (see Bennett-Levy, Richards, Farrand, & Christensen, 2010; White, 2008). The organisationally distinctive features of SC, however, include the delivery to large groups of participants (up to 100 in some services) in a 'night-class' style approach, which emphasises the lack of need for interaction with fellow attendees or facilitators. The high ratio of participants to facilitators makes SC an organisationally efficient treatment option for publically funded services required to treat large clinical populations (Kellett et al., 2007).

CBT has a robust evidence-base for the treatment of anxiety and depression problems (e.g., see Cuijpers et al., 2013; Hofmann & Smits, 2008). However, the high prevalence of these common mental health problems, coupled with the low availability and high cost of specialised psychotherapeutic treatments pose challenges to the accessibility of CBT in routine care (National Institute for Health and Care Excellence [NICE], 2011). In this regard, evidence-based high volume and low cost treatment options like SC could potentially help to meet the high demand for

depression and anxiety treatment in general primary care settings. The effectiveness of SC is supported by evidence from one controlled trial (White, Keenan, & Brooks, 1992), as well as a wide number of practice-based studies reviewed by Burns, Kellett and Donohoe (2016). Practice-based evidence suggests that SC participants on average experience a 50% reduction in anxiety and depression (Joice & Mercer, 2010; Wood, Kitchiner, & Bisson, 2005), although such studies do not include control group comparisons. In the most recent practice-based evaluation, Burns et al. (2016) reported a post-treatment recovery rate of 37% for SC participants and a dose-response relationship between the number of sessions attended and the likelihood of improvement.

Despite the growing evidence-base for SC, no multi-service studies have been conducted to date. The evidence base for SC is grounded in (often small) single-site studies, which have not enabled any cross-service comparisons. Therefore, important questions remain about the generalisability of treatment effects across organisations and teams, as well as the extent to which patient, clinician and group factors may explain the variability in clinical outcomes. Evidence from multi-service studies is necessary to assess the consistency of organisation, delivery, quality and outcomes of psychological healthcare (Weinberger et al., 2001). Multi-service studies are advantageous as they can provide large, diverse and externally-valid samples with sufficient statistical power to explore such questions (Gold & Dewa, 2005). This study sought to conduct the first multi-service evaluation of SC interventions routinely delivered in stepped care IAPT services. The study addressed the following research questions: (1) How consistent are clinical effects of SC across services? (2) Is clinical

effectiveness influenced by attendance rates? (3) Are clinical outcomes influenced by patient and/or group variables?

2. Method

2.1. Design and setting

This study was based on the analysis of historical routine practice data collected by 5 psychological therapy services linked to the Northern IAPT Practice Research Network (see Lucock et al., *submitted*). These services follow a stepped care model of treatment delivery (Clark et al., 2009; NICE, 2011). In this model, step 1 usually involves contact with a general medical practitioner (for assessment and consideration of options including pharmacological and psychological treatment), although some patients directly self-refer to psychological services. Step 2 includes low intensity psychoeducational interventions available in IAPT services including group and individual guided self-help as well as computerized CBT. Low intensity interventions are usually delivered across 6 to 8 sessions by trained psychological wellbeing practitioners and mental health nurses. Patients with more complex / severe disorders, and those who did not benefit from low intensity interventions, can access up to 20 sessions of formal (step 3) psychotherapeutic interventions. Step 3 interventions in this setting include CBT, interpersonal psychotherapy, counselling for depression, behavioural couples therapy and eye-movement desensitization and reprocessing (EMDR for post-traumatic stress disorder).

Together, the 5 participating services covered a geographical region including Cumbria, South and West Yorkshire in the north of England. The catchment area for these services included large, socio-economically and ethnically diverse cities (Sheffield, Leeds), as well as smaller towns (Barnsley,

Huddersfield), rural and semi-rural areas (in Cumbria, Kirklees, Calderdale). Available clinical pathway (treatments received), demographic and outcomes data (described below) were aggregated for all cases that accessed these services and were discharged from treatment between January 2013 and January 2015. Clinical collaborators at each service completed structured qualitative questionnaires to gather information on the delivery of SC interventions. Ethical approval to conduct the study was obtained from the North East - Newcastle & North Tyneside NHS research ethics committee (REC ref: 15/NE0062).

2.2. Measures and data sources

2.2.1. Clinical outcome measures

IAPT services are required to collect standardised patient-reported outcome measures on a session-to-session basis to monitor clinical progress. The GAD-7 is a seven-item measure developed to screen for anxiety disorders (Spitzer, Kroenke, Williams, & Löwe, 2006). Each item is rated on a 0 to 3 scale, yielding a total anxiety severity score between 0-21. A cut-off score ≥8 is recommended to identify the likely presence of a diagnosable anxiety disorder (Kroenke, Spitzer, Williams, Monahan, & Löwe, 2007). A change of ≥5 points defines reliable change on the GAD-7 (Richards & Borglin, 2011). The GAD-7 was the primary outcome measure in this study, given the focus of stress control interventions. The PHQ-9 is a nine-item screening tool for major depression (Kroenke, Spitzer, & Williams, 2001). Each item is also rated on a 0 to 3 scale, yielding a total depression severity score between 0-27. A cut-off ≥ 10 has been recommended to detect clinically significant depression symptoms (Kroenke, Spitzer, & Williams, 2001; Moriarty, Gilbody, McMillan, & Manea, 2015). A change of ≥6 points

defines reliable change on the PHQ-9 (Richards & Borglin, 2011). The Work and Social Adjustment Scale (WSAS) is a measure of functioning across five domains: work, home management, social leisure activities, private leisure activities, family and close relationships (Mundt, Marks, Shear, & Greist, 2002). Each item is rated on a scale of 0 (no impairment) to 8 (very severe impairment), rendering a total functional impairment score between 0–40, with no specific change (cut-off) parameter.

2.2.2. Secondary data

Clinical pathway data included information on treatments received (at steps 2 and 3 of the stepped care pathway), number of sessions attended and caseload variables which enabled the matching of each case to a specific SC group and a specific IAPT service. Demographic (age, gender, ethnicity, employment status, socioeconomic deprivation) and clinical characteristics (primary diagnosis, baseline severity in PHQ-9, GAD-7 and WSAS measures at assessment) were available for each case. Socioeconomic deprivation was derived by matching each patient's home postcode to the *English Index of Multiple Deprivation* (Department for Communities and Local Government, 2011), and categorising cases into quintile levels of deprivation (informed by Paddison et al., 2012).

2.3. Stress Control interventions: fidelity, delivery and organisation

All services based their interventions on the White (2008) SC model, which is structured as a six-session psycho-educational programme. Session 1 covers general information about stress and maintenance factors; session 2 covers relaxation skills and lifestyle changes; session 3 covers cognitive strategies to deal with automatic thoughts; session 4 covers problem solving

and activity scheduling; session 5 covers panic attack coping skills; session 6 covers sleep hygiene.

Three services made very minor modifications to content (e.g. number of power-point slides, terminology used). One service (service 'E') made substantial modifications to SC contents (e.g. reduced length of information, removed explanation of different anxiety disorders and information on relaxation skills) and abridged these into a shortened 5-session intervention. Modifications to contents and materials resulted in some differences in the length of sessions between services (ranging between 90-120 minutes). All services delivered SC in clinical (health centres) and community (public seminar and lecture rooms) venues and provided printed materials. Most services (4/5) allowed SC participants to be accompanied by friends or family if necessary. In all services SC was co-facilitated by 2 practitioners. SC facilitators were primarily psychological wellbeing practitioners, but 2 services also included other facilitators (nurses and psychological therapists). Inclusion criteria were generally broad and unrestrictive, although some services applied exclusions for people with severe depression/anxiety, dependent substance use, or diagnoses including OCD, social phobia and PTSD. All services had standard screening procedures to identify cases suitable for treatment in Primary Care, and 3 services enabled participants to self-book onto SC with minimal screening.

[Figure 1]

2.4. Participant characteristics and stepped care pathway

More than half of SC participants were female (63.1%), with a mean age of 42.94 (SD = 13.98; range: 16 – 89), and of White British ethnic background (92.6%). Most self-referred (71.4%), with the remainder referred by GPs (21.1%) or other professionals (7.5%). The most common primary presenting problems were mixed anxiety and depression (60.8%), GAD (19.7%) and depressive episode (11.1%). Mean baseline severity scores for the whole cohort were GAD-7 = 11.87 (SD = 5.33), PHQ-9 = 12.13 (SD = 6.02), WSAS = 14.82 (SD = 8.84). The mean number of group sessions attended was 4.26 (SD = 1.65; range = 1 – 9).

Figure 1 shows the flow of SC patients through the stepped care pathway. A total of 4,451 patients accessed 163 SC groups during the 2-year study period (range across services: 293 – 1675). Approximately 12.6% of cases receiving an intervention at step 2 received SC. SC groups had between 4 and 111 participants; mean = 48.77, SD = 27.42, median = 45. Based on prior research on low intensity interventions (Burns et al., 2016; Delgadillo et al., 2014; Firth, Barkham, Kellett, & Saxon, 2015), we applied a cut-off (≥4 sessions) to differentiate between SC completers and those who dropped out before receiving an adequate dose of SC. The treatment completion rate for SC was in the region of 70%. Approximately 15% accessed further treatment on completion of SC at steps 2 and 3, or were signposted to other services.

2.5. Data analysis

2.5.1. Benchmarking of clinical outcomes

Pre-post treatment effect sizes on the outcome measures for SC interventions (both in the whole sample and for each service) were calculated

with confidence intervals and critical values based on the equations proposed by Minami et al. (2008). Taking GAD-7 as the primary outcome measure (given the main focus on anxiety management in SC), between-service differences in effect sizes were compared using a forest plot and ANOVA. Effect sizes were compared to two benchmarks (pre-post Cohen's *d*); one benchmark derived from the only controlled trial of SC (White, Keenan, & Brooks, 1992) and the second benchmark derived from a meta-analysis of guided self-help interventions for anxiety and depression (Coull & Morris, 2011).

2.5.2. Dose-response analysis

Reliable and clinically significant improvement (RCSI) criteria (Jacobson & Truax, 1991) were applied to PHQ-9 and GAD-7 outcomes for each participant. To meet RCSI criteria, a patient with baseline scores in the symptomatic range (GAD-7 ≥ 8) should have sub-threshold post-treatment scores (GAD-7 < 8) and a pre-post change score greater than the reliable change index (reduction of at least 5 points in GAD-7). RCSI rates were then calculated for different clusters of participants attending the same number of SC sessions. This procedure enabled a bar chart to be plotted of RCSI rates for clusters of cases with the same SC treatment length, as well cumulative dose-response curves for each symptom measure.

2.5.3. Analysis of case-mix and group effects

Multilevel modelling (MLM) was applied to investigate whether SC outcomes were influenced by patient characteristics (case-mix), after controlling for differences between services and clustering within groups (*group effects*). Patients (level 1) were nested within SC groups (level 2) and

groups were nested within the 5 services (level 3). The post-treatment GAD-7 score was the independent variable, group was treated as a random effect and service was treated as a fixed effect. Service was treated as a fixed factor, because the small number of services precluded treating them as if they were randomly sampled from the wider population of IAPT services. Continuous variables were grand mean centred so coefficients can be interpreted in relation to the mean. This analysis was restricted to a subsample where each SC group had at least 5 participants (Total = 4,220 cases nested within 161 groups).

MLM was conducted in 4 steps. Model 1 was an unconditional model without any predictors other than the random effect for SC groups. Model 2 included fixed effects for the number of SC sessions attended and group size as a level-2 variable (i.e. an explanatory variable at group level). Model 3 added services as fixed effects in addition to model 2 variables. Finally, model 4 included fixed effects for case-mix variables: age, gender, ethnicity, employment status (employed vs. unemployed), index of multiple deprivation (IMD) quintile, baseline severity of symptoms (GAD-7, PHQ-9) and functional impairment (WSAS). This enabled the relative influence of group, service, and case-mix factors to be modelled. G*oodness-of-fit* for all models was assessed based on the Akaike information criterion (AIC), Bayesian information criterion (BIC) and -2 log likelihood statistics, and we tested if adding polynomial terms for continuous variables (sessions, age) improved model fit. An intra-class correlation coefficient (ICC) assessed the overall proportion of variance in GAD-7 outcomes attributable to the group level in each model (Raudenbush, 1993).

3. Results

[Figure 2]

3.1. Benchmarking of clinical effect sizes across services

Pre-post treatment effect sizes (Cohen's *d*) for the full sample were GAD-7 = 0.70 (95% CI: 0.66 to 0.73); PHQ-9 = 0.59 (95% CI: 0.56 to 0.62); WSAS = 0.47 (95% CI: 0.44 to 0.50). Effect sizes for cases that dropped out before receiving an adequate dose (attended <4 SC sessions) were considerably smaller (effect size range = 0.20 to 0.31). Figure 2 shows a forest plot of (GAD-7) pre-post effect sizes (and 95% confidence intervals with critical values) for each service, where the size of squares denotes differences in sample size, and the diamond shape represents the pooled effect size for all services. Four services had comparable effect sizes and this was not significantly different to the guided self-help efficacy benchmark (solid vertical line). The exception was in one service (service 'E') which had a significantly smaller effect size compared to other services, plus both benchmarks; F (4, 2933) = 4.29, $p < 0.01$. The pooled pre-post effect size for all services was significantly greater than the SC efficacy benchmark (dashed vertical line).

[Figure 3]

3.2. Clinical outcomes and attendance

On average, 41.6% of cases that initially scored in the clinical range (including completers and dropouts) met RCSI criteria by their last attended SC session (GAD-7 = 42.2%; PHQ-9 = 41.0%). Figure 3, shows a dose-response pattern suggesting that the greatest cumulative gains in recovery

were for those cases attending between 4-6 sessions. The curves superimposed onto the figure offer a visual representation of the cumulative percentage of cases that met RCSI criteria.

[Table 1]

3.3. *Multilevel modelling of service, group and patient variables*

A stepwise approach to multilevel modelling (MLM) was taken, as illustrated in Table 1. Model 1 with no covariates (i.e. variance components model) had a significant random effect estimate (Z = 36.841, p <.001), with an ICC value suggesting that 3.6% of variance in post-treatment anxiety scores was explained by variability between SC groups. On this basis, it was appropriate to account for the nested structure of the data in further analyses.

Model 2 (including covariates) suggested a curvilinear relationship (the sessions variable and its quadratic term were both significant predictors, p <.001) between the number of group sessions attended and post-treatment outcomes. This non-linear relationship is consistent with the dose-response curve in Figure 3. Group size (number of participants in each SC class) did not predict post-treatment anxiety scores (β = 0.005, SE = 0.005, p = .374).

Model 3 included services as covariates, confirming that services B (β = -1.567, SE = 0.419, p < .001) and D (β = -2.133, SE = 0.480, p < .001) tended to have lower post-treatment anxiety scores (better outcomes) compared to service E (which was the reference category).

Model 4 additionally included case-mix variables, confirming that higher post-treatment anxiety scores were found for cases in the most

socioeconomically deprived areas (IMD quintile 1, β = 0.720, SE = 0.338, p = .034) and those with higher baseline GAD-7 (β = 0.475, SE = 0.027, p < .001), PHQ-9 (β = 0.154, SE = 0.025, p < .001) and WSAS scores (β = 0.031, SE = 0.013, p = .020). Age, gender, ethnicity and employment status were not found to be statistically significant in this model (all had p < .05). Importantly, the service variable was no longer statistically significant (F (4, 1,842) = 0.993, p = .410) in model 4, suggesting that differences between services were fully explained by differences in group and case-mix variables.

Figure 4 displays a caterpillar plot of residuals (and 95% confidence intervals) for each of the 161 SC groups, ranking these from most to least effective in reducing anxiety (GAD-7) symptoms. The dashed reference line at 0 represents the average effect of SC interventions, and visually enables us to assess if each group's effects were equal to, above or below average. A negative residual denotes greater than average symptom reductions (better outcomes). The residuals are also colour coded according to service.

Univariate analyses (ANOVA) informed by the above MLM results confirmed that, compared to the other 4 services, patients in service E attended a lower mean number of SC sessions (F (4, 4804) = 28.483, p <.001), lived in more socioeconomically deprived areas (IMD; F (4, 4743) = 12.786, p <.001) and also had higher baseline anxiety (GAD-7; F (4, 3291) = 9.842, p <.001), depression (PHQ-9; F (4, 3256) = 10.836, p <.001) and functional impairment scores (WSAS; F (4, 3171) = 62.459, p <.001).

[Figure 4]

4. Discussion

4.1. Main findings

This practice research network study enabled a comprehensive evaluation of SC interventions in terms of effectiveness, efficiency and variability of outcomes between IAPT services. We found that large group SC interventions delivered in stepped care psychological services attain clinical effect sizes (pooled GAD-7 *d* = 0.70) comparable to those reported by the developers of the SC model (White, Keenan, & Brooks, 1992), and other controlled trials of guided self-help (GSH) for anxiety symptoms (Coull & Morris, 2011).

Although the SC treatment effects were fairly consistent across most services, there was evidence that one of the five participating services (service E) attained lower effect sizes which were in the moderate range (GAD-7 *d* = 0.48). SC delivered at this service deviated from the standard treatment protocol, with psychoeducational materials condensed into a shortened 5-session group programme. Compared to other participating services, patients at this service were more socioeconomically disadvantaged and had higher levels of symptom severity and functional impairment. An adequately powered multilevel modelling analysis demonstrated that outcome differences between services were largely explained by these differences in therapy length and case-mix variables.

The above finding demonstrates that the way in which evidence-based interventions are adopted can influence their effectiveness in routine practice. The *implementation science* literature suggests that the successful dissemination of novel approaches into clinical care can be influenced by internal (e.g., organisational structures, culture, priorities, readiness) and external (e.g., funding, policy influences) factors (Aarons, Hurlburt, &

Horwitz, 2011; Meyers, Durlack, & Wandersman, 2012). English stepped care services are operating within a policy context where national targets require them to considerably increase the number of patients accessing psychological care and also reduce waiting lists (Department of Health, 2014). Such external conditions are likely to influence the way in which evidence-based treatments are adapted and implemented; though other internal factors in service E may have also influenced their decision to modify the SC intervention (unlike other services that are under similar external pressures).

Our findings crucially underline the importance of maintaining fidelity to the evidence-base when research-based interventions are disseminated into routine care. Meyers et al. (2012) propose that the process of implementation requires an explicit assessment of how innovations may need to be adapted to a specific practice setting, coupled with a process evaluation and the establishment of feedback mechanisms. SC has been widely disseminated across numerous services in England, some of which included process evaluations (e.g., Burns et al., 2016). However, as we have seen, the successful implementation in one service does not necessarily guarantee generalisability elsewhere. Ideally, services adopting (and adapting) any evidenced–based interventions should endeavour to establish a data-based feedback and clinical audit cycle as part of their implementation plans. The benchmarking method illustrated in this study could be used to support such implementation and evaluation efforts in similar contexts.

Our finding that some process and patient variables moderate the effectiveness of psychoeducational CBT is consistent with the wider literature. Two prior studies using data from different IAPT services

concluded that the effectiveness of low intensity interventions is maximised between 4 to 6 sessions; additional sessions after this point rarely lead to better outcomes (Delgadillo et al., 2014; Firth et al., 2015). In fact, Burns et al. (2016) found that SC patients who had additional concurrent treatment at step 2 (low intensity) did not attain superior outcomes to those who simply attended SC. Our dose-response analysis adds further evidence for this 'optimal dose of psychoeducation' pattern. Previous studies have also demonstrated that patient-factors such as higher baseline severity of depression, severe functional impairment, socioeconomic poverty and unemployment predict poorer outcomes in low intensity psychological interventions (Delgadillo, Asaria, Ali, & Gilbody, *in press*; Delgadillo, Moreea, & Lutz, 2016; Firth et al., 2015).

To our knowledge, this is the first study to investigate the extent to which psychoeducation outcomes are influenced by differences between groups (e.g., *group effects*). Differences between groups were found to explain up to 3.6% of variance in post-treatment outcomes. This estimate is smaller by comparison to the general influence of *therapist effects* in formal 'high intensity' psychotherapy (between 5% and 10%; Baldwin & Imel, 2013), but closer to those of low intensity interventions delivered in IAPT services (between 1% and 9%; Ali et al., 2014; Green et al., 2014; Firth et al., 2015). Though the group estimate is small, it is remarkable that a highly standardised and manualised psychoeducational intervention with minimal therapist-patient interaction should yield between-group differences. It is plausible that variability in facilitators' competence and delivery may partly explain these group effects (Burlingame, Strauss, & Joyce, 2013). In this regard, future research could focus on the development of methods to measure and to enhance facilitator competency. It is also possible that other

factors related to the group context (e.g., self-referral versus professional referral, accessibility of venues, the extent to which the atmosphere is perceived as welcoming and non-threatening) may play into group effects, though further research is necessary to verify this.

This study explored the potential influence of one such contextual variable: group size, which was not associated with clinical effects. Patients in SC groups as large as 100 participants were equally likely to benefit from the intervention as those in smaller groups. The size of SC groups may have a normalising effect for participants and further research on SC mechanisms of change is needed. The finding regarding group size strengthens the argument that high volume SC classes are likely to be an organisationally efficient treatment option. The clinical caveat to this assertion, however, is that some patients with particular characteristics are much less likely to benefit from SC and tend to drop out of care early on. We also noted that effect sizes for PHQ-9 (d = .59) and WSAS (d = .47) were more modest compared to those for anxiety outcomes measured using GAD-7 (d = .70). This suggests that patients with more severe depression and functional impairment derive less benefit from SC interventions; which is consistent with prior outcome-prediction studies of low intensity interventions (Delgadillo et al., 2016).

4.2. Limitations

The pre-post treatment effect sizes described in this study offer a general estimate of the 'real world' effectiveness of SC interventions delivered in routine stepped care services. As a naturalistic cohort study, these effect sizes are not assessed relative to control groups, and therefore it is possible that regression to the mean (i.e., natural fluctuations in mental health

symptoms due to the passage of time) may have partly accounted for some of the reported effects. Furthermore, our uncontrolled data do not allow us to disentangle specific SC treatment effects from effects that may be due to general contact with healthcare practitioners and other patients in a group-based setting. The small number of services clustered in Northern England may not necessarily be representative of populations in other regions of the country or internationally, so further replication in other regions would help to establish the generalisability of SC effects with greater certainty. Although we were able to link individual patients to specific SC groups, the available data did not enable us to match groups to pairs of facilitators. Therefore, it was not possible to examine the influence of specific facilitators, who may possibly vary in fidelity, competence and credibility. Standardised SC fidelity measures or checklists were not available or routinely collected in these services, so we relied on self-reported qualitative data to determine the extent to which SC treatments were standardised across services. Furthermore, an important limitation of this study is that we were only able to examine short-term outcomes, since post-treatment follow-up data were not available. Research on the durability of clinical effects over longer follow-up periods is necessary to support the evidence-base for large group psychoeducation.

4.3. Clinical implications

We propose five key points that may maximise the effectiveness of SC within the context of stepped care. (1) Patients should be made aware of alternative treatment options and should be able to make an informed choice. This may be particularly important for those with known disadvantages that may hinder their likelihood of benefit from SC (severe

depression and anxiety symptoms, severe functional impairment, marked socioeconomic deprivation). (2) SC classes should be delivered with fidelity to the original 6-session treatment protocol. (3) SC participants should be encouraged and supported to attend all 6 sessions and information regarding the benefits of attendance could be included at session 1. (4) SC participants who do not show signs of improvement after having an 'adequate dose' (4 to 6 sessions) should be offered more personalised and/or intensive treatment options. (5) Post-treatment follow up may be a worthwhile addition to SC, for instance by planning 'booster sessions' as in traditional CBT interventions.

4.4 Conclusions

This study has illustrated that large group psychoeducational CBT is an important component of the suite of interventions offered at the early stages of the stepped care model in IAPT services. The effectiveness of SC in routine practice appears to be comparable to that of other brief interventions, including individual guided self-help and computerized CBT. The delivery ratio (up to 100 participants per 2 facilitators) and minimal need for clinician-patient contact are unique aspects that enable SC to considerably enhance access to psychoeducational support at low cost. Socioeconomic context, initial severity and subsequent attendance are all important predictors of outcome. There is also clear evidence of variability of outcomes between groups and hence a 'group effect', which suggests that services should attend to the facilitators' competence and fidelity of delivery of psychoeducational materials. Perhaps the most important finding is that decisions to shorten and adapt extant evidence-based practice can have an unforeseen impact on patient outcomes.

Acknowledgements

With thanks to Moira Bowler, Saim Mahmood, Corinne Mallinson, Cath Johnston, Alan Archer and Sam Temple for support with the acquisition and preparation of datasets. Thanks also to Victoria Adam, Laurence Gregory, Katherine Lofthouse and Zina Muftin for contributing to the study steering group.

The Stress Control Project was supported by NHS Research Capability Funding from the West Yorkshire Clinical Commissioning Groups, United Kingdom. The study was conducted in collaboration with clinical and academic collaborators affiliated to the Northern IAPT Practice Research Network.

References

Aarons, G. A., Hurlburt, M., & Horwitz, S. M. (2011). Advancing a conceptual model of evidence-based practice implementation in public service sectors. *Administration and Policy in Mental Health and Mental Health Services Research*, *38*(1), 4–23.

Ali, S., Littlewood, E., McMillan, D., Delgadillo, J., Miranda, A., Croudace, T., & Gilbody, S. (2014). Heterogeneity in patient-reported outcomes following low-intensity mental health interventions: A multilevel analysis. *PLoS ONE*, *9*(9), e99658. doi:10.1371/journal.pone.0099658

Baldwin, S.A., & Imel, Z.E. (2013). Therapist effects: Findings and methods. In M. J. Lambert (Ed.), *Bergin and Garfield's handbook of psychotherapy and behavior change (6th ed.)*. New Jersey: John Wiley & Sons Inc.

Bennett-Levy, J., Richards, D., Farrand, P., Christensen, H. (2010). *Oxford guide to low intensity CBT interventions*. Oxford: Oxford University Press.

Burlingame, G.M., Strauss, B., & Joyce, A.S. (2013). Change mechanisms and effectiveness of small group treatments. In: M. J. Lambert (Ed.), *Bergin and Garfield's handbook of psychotherapy and behavior change (6th ed.)*. New Jersey: John Wiley & Sons Inc.

Burns, P., Kellett, S., & Donohoe, G. (2016). "Stress Control" as a large group psychoeducational intervention at Step 2 of IAPT services:

Acceptability of the approach and moderators of effectiveness. *Behavioural and Cognitive Psychotherapy*, 44(4), 431-443.

Clark D. M. (2011). Implementing NICE guidelines for the psychological treatment of depression and anxiety disorders: The IAPT experience. *International Review of Psychiatry*, *23*(4), 318–327.

Clark, D. M., Layard, R., Smithies, R., Richards, D. A., Suckling, R., & Wright, B. (2009). Improving access to psychological therapy: initial evaluation of two UK demonstration sites. *Behaviour Research and Therapy*, *47*, 910–920.

Coull, G., & Morris, P.G. (2011). The clinical effectiveness of CBT-based guided self-help interventions for anxiety and depressive disorders: a systematic review. *Psychological Medicine*, *41*, 2239–2252.

Cuijpers, P., Berking, M., Andersson, G., Quigley, L., Kleiboer, A., & Dobson, K. S. (2013). A Meta-Analysis of Cognitive-Behavioural Therapy for Adult Depression, Alone and in Comparison With Other Treatments. Canadian *Journal of Psychiatry, 58*(7), 376–385.

Delgadillo, J., Asaria, M., Ali, S., & Gilbody, S. (*in press*). On poverty, politics and psychology: The socioeconomic gradient of mental healthcare utilisation and outcomes. *The British Journal of Psychiatry*.
doi: 10.1192/bjp.bp.115.171017

Delgadillo, J., McMillan, D., Lucock, M., Leach, C., Ali, S., & Gilbody, S. (2014). Early changes, attrition, and dose-response in low intensity psychological interventions. *British Journal of Clinical Psychology*, *53*, 114-130.

Delgadillo, J., Moreea, O., Lutz, W. (2016). Different people respond differently to therapy: A demonstration using patient profiling and risk stratification. *Behaviour Research and Therapy*, *79*, 15-22.

Department for Communities and Local Government. (2011). *The English Indices of Deprivation 2010*. London: Author.

Department of Health. (2014). Closing the Gap: Priorities for Essential Change in Mental Health. Department of Health, 2014. Available at: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/281250/Closing_the_gap_V2_-_17_Feb_2014.pdf

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

Gold, J.L., & Dewa, C.S. (2005). Institutional review boards and multisite studies in health services research: is there a better way? *Health Services Research*, *40*(1), 291-307.

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.

Hofmann, S. G., & Smits, J. A. J. (2008). Cognitive-Behavioral Therapy for Adult Anxiety Disorders: A Meta-Analysis of Randomized Placebo-Controlled Trials. *Journal of Clinical Psychiatry*, *69*, 621-632.

Jacobson, N., & Truax, P. (1991). Clinical significance: A statistical approach to defining meaningful change in psychotherapy research. *Journal of Consulting and Clinical Psychology*, *59*, 12–19.

Joice, A., & Mercer, S. (2010). An evaluation of the impact of a large group psycho-educational programme (Stress Control) on patient outcome: does empathy make a difference? *The Cognitive Behaviour Therapist*, *3*, 1–17.

Kellett, S., Clarke, S., & Matthews, L. (2007). Session impact and outcome in group psycho-educative cognitive behavioural therapy. *Behavioural and Cognitive Psychotherapy*, *35*, 335–342.

Kroenke, K., Spitzer, R. L., Williams, J. B. W., Monahan, P. O., & Löwe, B. (2007). Anxiety disorders in primary care: Prevalence, impairment, comorbidity, and detection. *Annals of Internal Medicine*, *146*(5), 317–325.

Lucock, M., Delgadillo, J., Barkham, M., Donohoe, G., Kellett, S., McMillan, D., Mullaney, S., Sainty, A., Saxon, D., & Thwaites, R. (*submitted*). The Northern IAPT Practice Research Network: Feasibility, Processes and Learning.

Meyers, D. C., Durlack, J. A., & Wandersman, A. (2012). The quality implementation framework: A synthesis of critical steps in the implementation process. *American Journal of Community Psychology, 50*(3–4), 462–480.

Minami, T., Serlin, R. C., Wampold, B. E., Kircher, J. C., & Brown, G. S. (2008). Using clinical trials to benchmark effects produced in clinical practice. *Quality and Quantity*, *42*, 513–525.

Moriarty, A.S., Gilbody, S., McMillan, D., Manea, L. (2015). Screening and case finding for major depressive disorder using the Patient Health Questionnaire (PHQ-9): a meta-analysis. *General Hospital Psychiatry*, *37*(6), 567–576.

Mundt, J. C., Mark, I. M., Shear, M. K., & Griest, J. M. (2002). The Work and Social Adjustment Scale: a simple measure of impairment in functioning. *The British Journal of Psychiatry*, *180*(5), 461-464.

National Institute for Health and Clinical Excellence (2011). *Common Mental Health Disorders: identification and pathways to care*. London: National Collaborating Centre for Mental Health.

Paddison, C., Elliott, M., Parker, R., Staetsky, L., Lyratzopoulos, G., Campbell, J. L., & Roland, M. (2012). Should measures of patient experience in primary care be adjusted for case mix? Evidence from the English General Practice Patient Survey. *BMJ Quality and Safety*, *21*, 634–640.

Raudenbush, S. W. (1993). Hierarchical linear models and experimental design. In L. Edwards (Ed.), *Applied analysis of variance in behavioral science* (p. 459 - 496). New York: Marcel Dekker.

Richards, D. A., & Borglin, G. (2011). Implementation of psychological therapies for anxiety and depression in routine practice: Two year prospective cohort study. *Journal of Affective Disorders*, *133*, 51–60.

Spitzer, R., Kroenke, K., Williams, J. B. W., & Lowe, B. (2006). A brief measure for assessing generalized anxiety disorder: The GAD-7. *Archives of Internal Medicine*, *166*(10), 1092–1097.

Weinberger, M., Oddone, E.Z., Henderson, W.G., Smith, D.M., Huey, J., Giobbie-Hurder, A., Feusner, J.R. (2001). Multisite randomized controlled trials in health services research: scientific and operational issues. *Medical Care*, *39*, 627–634.

White, J. (2008). CBT and the challenge of primary care: developing effective, efficient, equitable, acceptable and accessible services for common mental health problems. *Journal of Public Mental Health*, *7*, 32–41.

White, J., & Keenan, M. (1990). “Stress Control”: a pilot study of large group therapy for generalized anxiety disorder. *Behavioural Psychotherapy*, *18*, 143–146.

White, J., Keenan, M., & Brooks, N. (1992). ‘Stress Control’: A controlled comparative investigation of large group therapy for Generalized Anxiety Disorder. *Behavioural Psychotherapy*, *20*(2), 97-114.

Wood, S., Kitchiner, N. J., & Bisson, J. I. (2005). Experiencing and implementing an adult educational approach to treating anxiety disorder. *Journal of Psychiatric and Mental Health Nursing*, *12*, 95–99.

Table 1.

Multilevel modelling of the influence of group, service and case-mix factors on post-treatment anxiety scores (GAD-7)

*Variance components			Summary of main effects				
Residual variance estimate	Random effect variance estimate	Group effects (ICC)	Variable	β	SE	*p*	*95% CI*
MODEL 1 (AIC = 17,970.147; BIC = 17,982.064; -2 log likelihood = 17,966.143)							
30.091	1.133	3.6%					
MODEL 2 (AIC = 17,618.588; BIC = 17,630.502; -2 log likelihood = 17,614.584)							
26.797	0.471	1.7%					
			intercept	8.429	0.137	<.001	8.160, 8.698
			sessions	-2.232	0.278	<.001	-2.777, -1.687
			sessions^2	0.152	0.037	<.001	0.079, 0.225
MODEL 3 (AIC = 17,592.921; BIC = 17,604.833; -2 log likelihood = 17,588.917)							
26.739	0.264	1.0%					
			intercept	9.397	0.385	<.001	8.643, 10.151
			sessions	-2.350	0.278	<.001	-2.895, -1.805
			sessions^2	0.172	0.037	<.001	0.100, 0.245
			**service = B vs E	-1.567	0.419	<.001	-2.388, -0.746
			**service = D vs E	-2.133	0.480	<.001	-3.075, -1.192
MODEL 4 (AIC = 10,365.767; BIC = 10,376.798; -2 log likelihood = 10,361.760)							
14.879	0.089	0.6%					
			intercept	10.037	1.551	<.001	6.995, 13.080
			sessions	-1.958	0.253	<.001	-2.455, -1.461
			sessions^2	0.146	0.034	<.001	0.080, 0.212
			***IMD quintile = 1 vs 5	0.720	0.338	.034	0.056, 1.384
			baseline GAD-7	0.475	0.027	<.001	0.423, 0.527
			baseline PHQ-9	0.154	0.025	<.001	0.105, 0.203
			baseline WSAS	0.031	0.013	.020	0.005, 0.057

Notes: All continuous variables are mean centred; ICC = intra-class correlation coefficient; β = fixed coefficients; SE = standard error; CI = confidence intervals; AIC = Aikake (corrected) information criterion; BIC = Bayesian information criterion; sessions^2 = quadratic term for sessions attended; * Z-tests for all variance estimates were significant at p <.01; **service = 5-level variable with service E as the reference category; ***IMD = 5-level variable representing quintiles of socioeconomic deprivation with quintile 5 (least deprived) as reference category; non-significant fixed effects are excluded from the table (stress control group size, age, gender, ethnicity, employment status)

Figure 1. Stepped care pathway for stress control participants

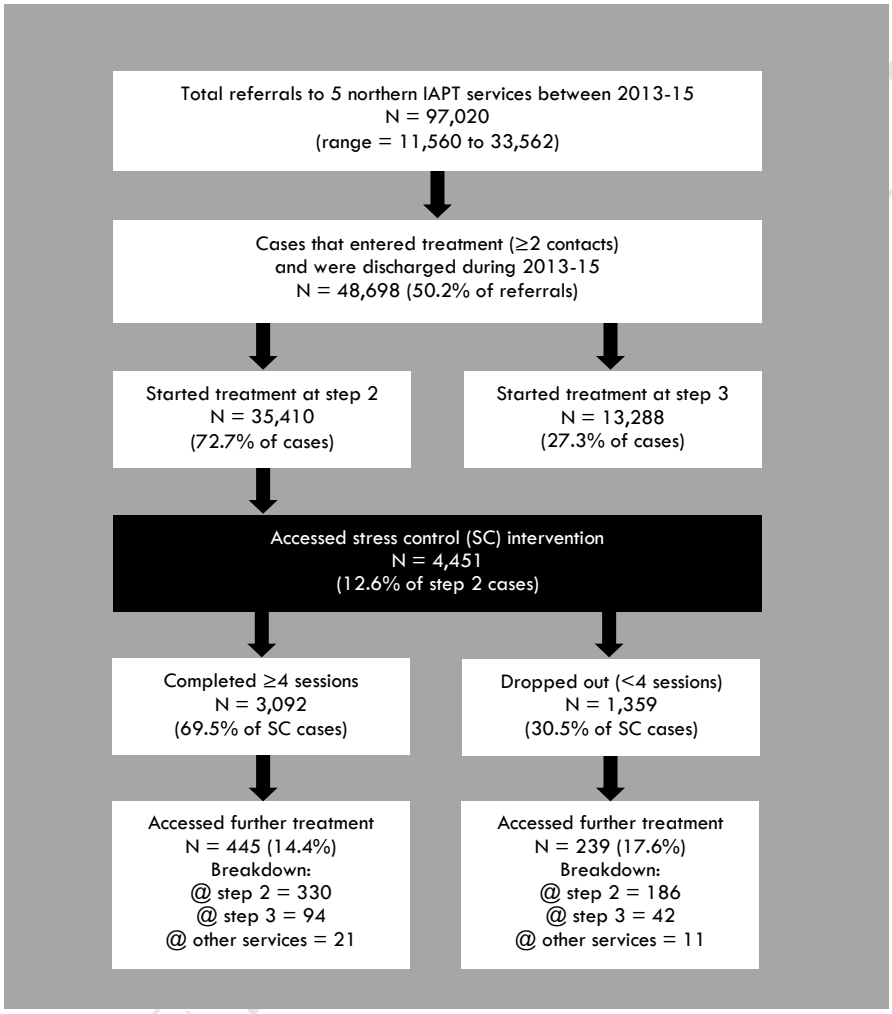


Figure 2. Benchmarking analysis of SC interventions across 5 IAPT services

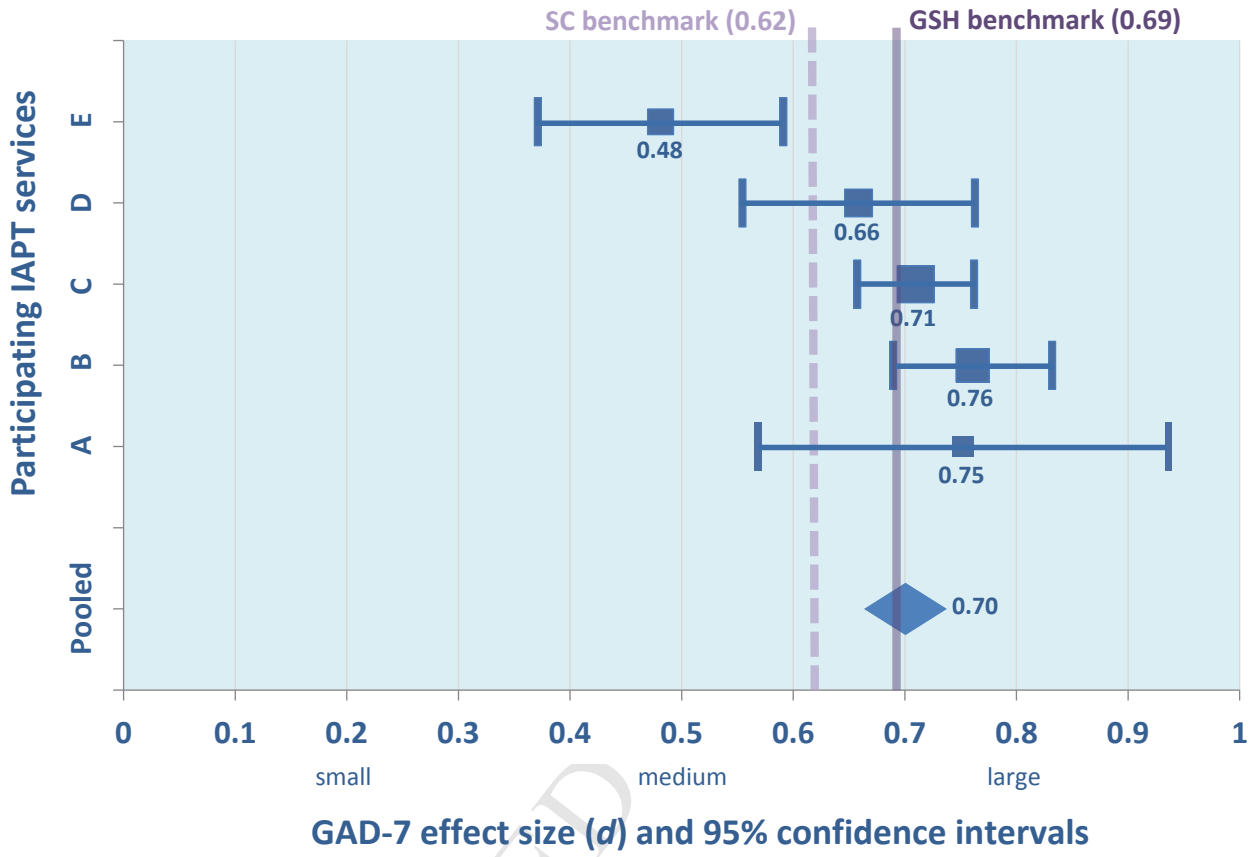


Figure 3. Dose-response in stress control interventions

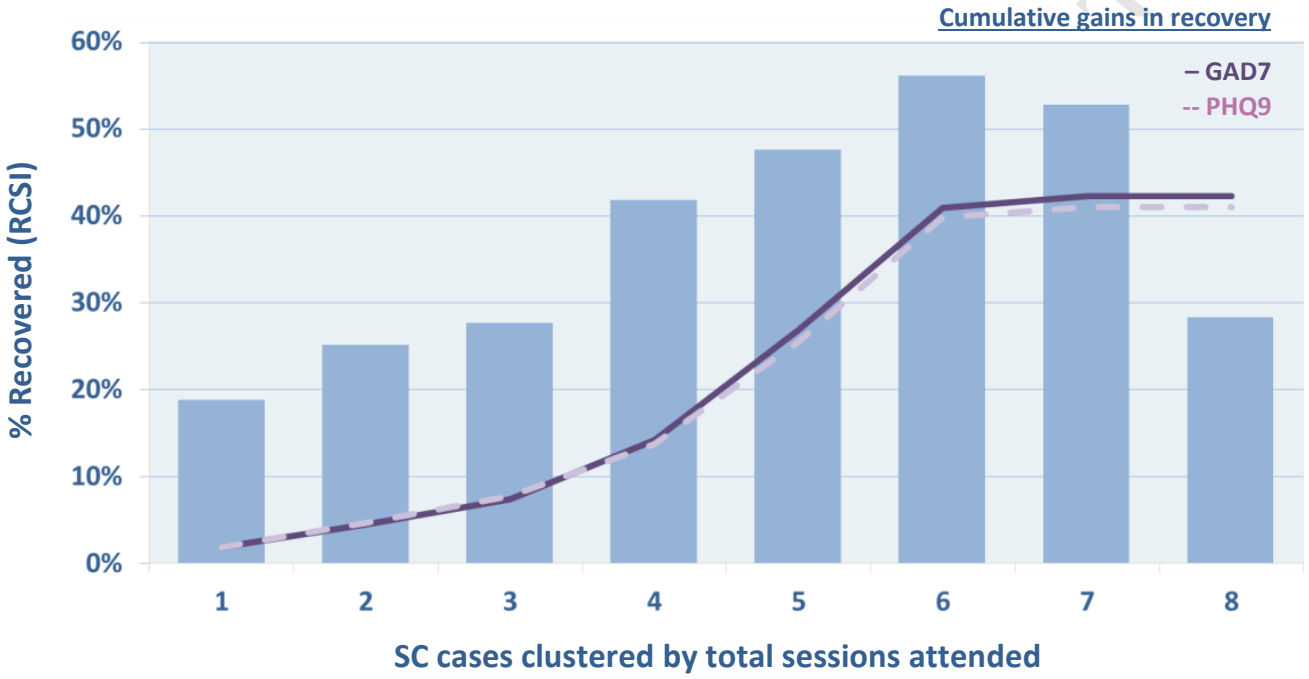
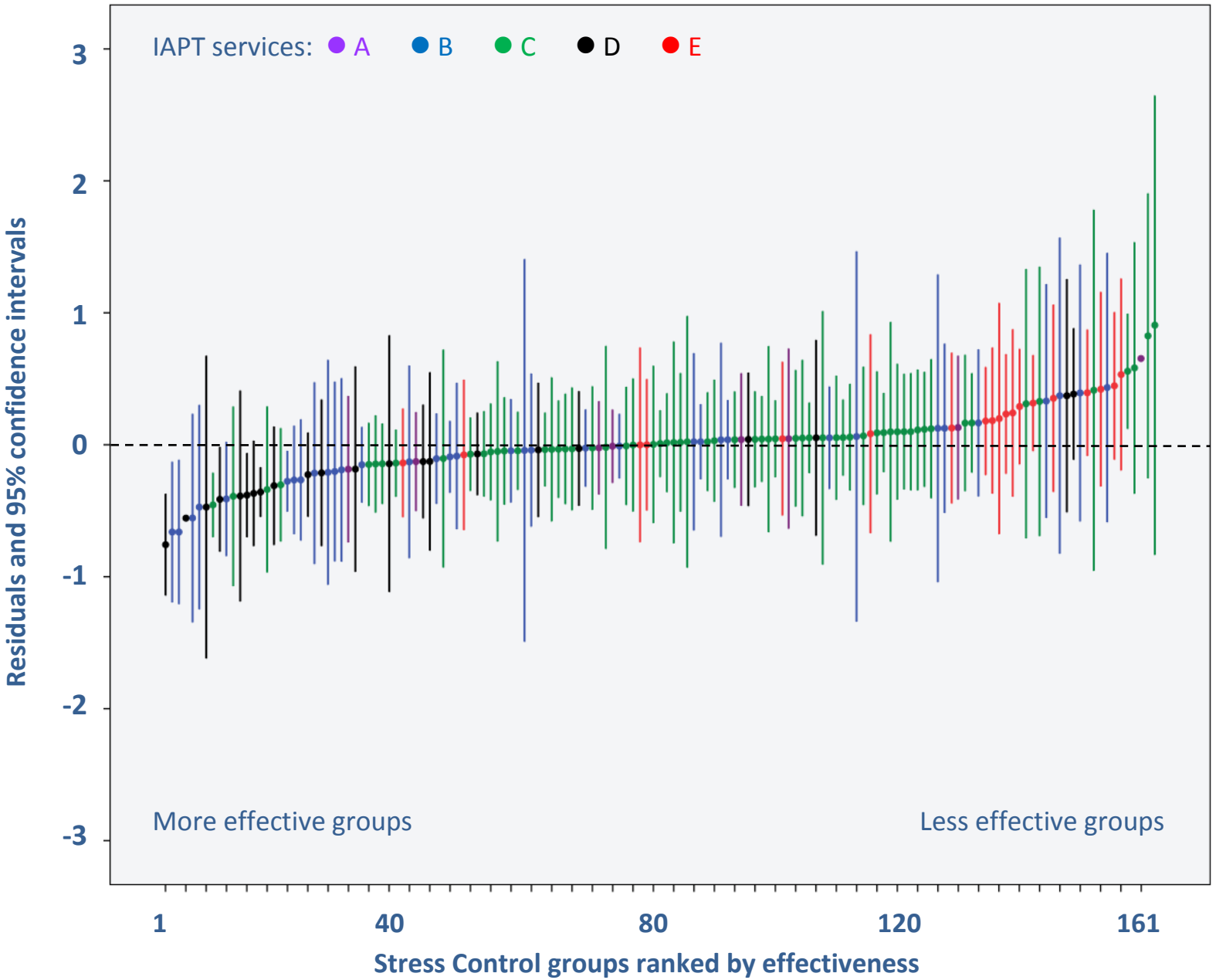


Figure 4. Caterpillar plot: variability in GAD-7 outcomes across groups



Highlights

- We analysed clinical outcomes data for 4,451 participants in 163 psychoeducational groups delivered by 5 services
- Pooled clinical effect sizes were similar to benchmarks from a meta-analysis of guided self-help trials
- Results indicate that differences between groups explain up to 3.6 % of variability in outcomes
- Outcome differences between services were largely explained by group and patient variables