Behavioural and Cognitive Psychotherapy, 2021, **49**, 76–90
doi:10.1017/S1352465820000624

MAIN

A service evaluation of a group mindfulness-based intervention for distressing voices: how do findings from a randomized controlled trial compare with routine clinical practice?

Anna-Marie Jones[1], Clara Strauss[1,2] and Mark Hayward[1,2,*]

[1]R&D Department, Sussex Education Centre, Sussex Partnership NHS Foundation Trust, Nevill Avenue, Hove BN3 7HZ, UK and [2]School of Psychology, University of Sussex, Pevensey Building, Falmer BN1 9QH, UK
*Corresponding author. Email: m.i.hayward@sussex.ac.uk

(Received 7 February 2020; revised 15 June 2020; accepted 16 July 2020; first published online 25 September 2020)

Abstract

Background: Person-based cognitive therapy (PBCT) was developed as a treatment for psychosis. The effectiveness of group PBCT was examined in the Mindfulness for Voices (M4V) randomized controlled trial and generated promising results. Group PBCT was implemented as a trans-diagnostic treatment for distressing voices within the Sussex Voices Clinic (SVC), a specialist secondary care mental health service.

Aim: To conduct a service evaluation of engagement, outcomes and cost of group PBCT within SVC, and to compare engagement and outcomes from routine practice with the M4V trial. Secondary aims were to explore predictors of levels of engagement and change in group PBCT.

Method: Service level data from 95 SVC patients were evaluated. Descriptive statistics, hypothesis tests and linear regression models were used. The primary clinical outcome was voice-related distress. Engagement levels and pre–post effect sizes were estimated; associated predictors were explored.

Results: Fifty-nine per cent of patients completed group PBCT within SVC, compared with 72% within M4V. Completers within SVC had lower baseline depression scores compared with non-completers. There were significant improvements in voice-related distress (Cohen's $d = -0.47$; $p = 0.001$), subjective recovery (Cohen's $d = 0.35$; $p = 0.001$) and depression (Cohen's $d = -0.20$; $p = 0.044$); these outcomes were comparable to M4V. Higher baseline subjective recovery and lower depression both predicted improvement in voice-related distress. Therapy within SVC cost an average of £214 per patient.

Conclusion: PBCT groups can be delivered trans-diagnostically in routine clinical practice. Engagement was lower when compared with an RCT, but outcomes were comparable. The low level of resources involved suggests that group PBCT can offer value for money.

Keywords: auditory verbal hallucinations; cognitive therapy; group therapy; mindfulness; psychosis; trans-diagnostic; voice-hearing

Introduction

Person-based cognitive therapy (PBCT; Chadwick, 2006) integrates cognitive behavioural therapy for psychosis (CBTp) with a mindfulness-based approach and was developed specifically for people experiencing distressing psychosis. PBCT explicitly targets three distinct sources of distress: persecutory delusions and beliefs about voices ('symptomatic meaning'); self-defeating reactions to psychotic symptoms (e.g. experiential avoidance, fighting with voices and paranoid rumination); and core beliefs (schemata) that define the self as negative and fixed (Chadwick, 2006). The therapy combines guided discovery, behavioural experiments, a focus on positive behaviour change, mindfulness practice and a strong experiential focus. The effectiveness of PBCT delivered in a group format was examined in the Mindfulness for Voices (M4V) trial (registration ISRCTN74054823). This was the first randomized controlled trial (RCT) to evaluate group PBCT in people with a diagnosis of psychosis. It was designed as a multi-site, assessor-blinded, pragmatic study with 1:1 allocation to either treatment-as-usual or group PBCT. Findings from the M4V trial indicated that PBCT, offered in a group format and delivered over 12 sessions, was a promising intervention for reducing levels of voice-related distress and depression, and promoting recovery (Chadwick *et al.*, 2016). Following the successful completion and encouraging findings of the M4V trial, group PBCT was implemented within routine clinical practice through the Sussex Voices Clinic (SVC). SVC is a specialist outpatient service within the secondary care services of the National Health Service (NHS) in the United Kingdom (UK) and offers evidence-based interventions to patients distressed by hearing voices, irrespective of diagnosis.

In the context of the limited availability of evidence-based psychological therapy for psychotic experiences, including distressing voices (Royal College of Psychiatrists, 2018), PBCT groups may add value as they have the potential to be less resource intensive. Within the M4V trial, each participant who completed a PBCT group received, on average, seven hours of therapist time. This compares favourably with the 16 hours recommended by NICE (National Institute for Health and Care Excellence, 2014) to generate similar outcomes from individual therapy. Furthermore, groups can generate the sense of ‘universality’ (Goodliffe *et al.*, 2010) and ‘connectedness’ (Payne *et al.*, 2017) that are valued by patients distressed by hearing voices. Mindfulness-based therapy groups for voice hearers are also viewed as helpful by staff and patients (Morera *et al.*, 2015).

In order to maximize both benefits for patients and added value, patients need to attend group PBCT sessions regularly. However, some patients have to overcome many barriers and challenges when trying to complete the 12 sessions (McHale *et al.*, 2018). A minimum effective dose of group PBCT, or therapy completion, is considered to be attendance to at least eight out of 12 sessions (Chadwick *et al.*, 2016). Little is known about which patients are at risk of disengagement from group PBCT therapy or of not benefiting. Having an understanding of which patient characteristics and baseline factors are associated with these risks will help clinicians to make informed decisions about who should be offered the group therapy and when (Paulik *et al.*, 2018). Importantly, this will help to maximize the likelihood of patient benefit whilst helping services to optimize the use of limited resources.

To our knowledge this is the first study to evaluate group PBCT for people who hear voices in a real-world setting. This service evaluation reports on the experience of offering PBCT groups within SVC, with reference to the resources and costs required to deliver the groups within routine clinical practice. The key questions that underpinned this service evaluation were: when delivered in routine clinical practice (1a) how well do patients engage with PBCT groups?, (1b) what clinical outcomes are experienced by patients who complete PBCT therapy? and (1c) how do these findings contrast with our experience of offering PBCT groups within the M4V trial?; (2) what resources are needed to deliver PBCT and what are the associated costs?; (3) can engagement levels or clinical change be predicted by any baseline patient characteristics? We hypothesize that both engagement levels and clinical outcomes will be lower in the naturalistic setting of SVC compared with the results from the M4V trial. Meanwhile, whilst our assessment of predictors is exploratory, we hypothesize that negative affect and baseline voice-related distress levels may be instrumental factors. This would be in line with findings from a study that assessed predictors of engagement with, and outcomes from, a brief individual intervention within SVC (Paulik *et al.*, 2018).

Method

Study design

This was a study of engagement, pre–post clinical change, potential predictors and delivery costs for PBCT groups delivered in a naturalistic and uncontrolled setting of routine clinical care at a single NHS site. Assessment measures were administered by clinic assistants who were not involved in therapy delivery. Throughout their attendance at SVC, patients received treatment-as-usual from their mental health teams. This consisted of both regular outpatient appointments with a consultant psychiatrist and their care coordinator, and psychotropic medication. As this study was a service evaluation of routine clinical practice, NHS Research Ethics Committee approval was not required (UK Policy Framework for Health and Social Care Research; Department of Health, 2017). This service evaluation was registered with an NHS audit department (dated 26 August 2015) who advised that informed patient consent was not necessary. All data in this evaluation have been anonymized.

Patients

SVC is a trans-diagnostic outpatient service in secondary care within a single NHS Mental Health Trust in Sussex, UK. Before November 2016, the inclusion criteria for SVC was as follows: (a) a score of 4 or above on the ‘hallucinatory behaviour’ item on the Positive and Negative Symptom Scale (PANSS; Kay et al., 1987) and (b) a score of at least 3 on one of the distress items (‘intensity of distress’ and ‘amount of distress’) of the Psychotic Symptom Rating Scales–Auditory Hallucinations (PSYRATS-AH; Haddock *et al.*, 1999). In November 2016, the inclusion criteria changed to a score of at least 8 on the Negative Impact scale of the Hamilton Program for Schizophrenic Voices Questionnaire (HPSVQ; Van Lieshout and Goldberg, 2007). SVC moved from the PSYRATS-AH, which is observer-rated to the HPSVQ, a self-report measure, in order to reduce the burden on patients and clinic assistants; the HPSVQ has additional advantages through its convenience, acceptability to patients and ability to accurately depict subjective factors (Kim *et al.*, 2010). The factor structures of the PSYRATS-AH and HPSVQ have also been shown to discriminate emotional and physical items in a similar way (Kim *et al.*, 2010).

When patients were referred to SVC, they initially received a baseline assessment. If they met the eligibility criteria, they were offered four sessions of individual Coping Strategy Enhancement (CSE; Hayward *et al.*, 2018) as a ‘Level 1’ intervention followed by a post-Level 1 assessment. If they remained eligible, the patient was offered a ‘Level 2’ intervention of 12 sessions of PBCT group therapy. If the post-Level 1 assessment suggested that engagement with a group intervention would not be possible (due to either interpersonal sensitivity or practical issues), a course of individual therapy was offered as an alternative. The individual therapies were offered over 8 (guided self-help CBT; Hazell *et al.*, 2018a,b) or 16 (Relating Therapy; Hayward *et al.*, 2017) sessions. The Level 2 intervention was followed by a final post-Level 2 assessment. Patients included in this evaluation are those who were offered and initially accepted the offer of the PBCT group intervention at Level 2. Between November 2014 and March 2019, a total of 95 patients were invited to attend one of 13 PBCT groups; 53 patients were offered individual therapy.

Clinical measures

Assessment measures were collected by a clinic assistant not involved in the delivery of therapy (to reduce potential bias) within 4 weeks of the group starting (pre-PBCT) and within 4 weeks of its completion (post-PBCT). The patient’s psychiatrist confirmed patient diagnosis, and the clinic assistant collected demographic information when patients first entered SVC. The following measures were administered to assess clinical outcomes.

Primary outcomes

Psychotic Symptom Rating Scales – Auditory Hallucinations (*PSYRATS-AH*) (Haddock *et al.*, 1999). This is an 11-item semi-structured interview designed to measure the different dimensions of auditory hallucinations. Factor analysis has shown that the scale has four dimensions (Woodward *et al.*, 2014): *distress* (negative content, distress and control); *frequency* (frequency, duration and disruption); *attribution* (location and origin of voices); and *loudness* (loudness item only). Haddock *et al.* (1999) reported inter-rater reliability scores for the five individual distress items ranging from 0.8 to 1.00, and Woodward *et al.* (2014) reported a high intraclass correlation (ICC) of 0.93 for the distress scale. The distress items were as follows: Q6, Amount of negative content of voices; Q7, Degree of negative content; Q8, Amount of distress; Q9, Intensity of distress; and Q11, Controllability of voices. Each item was scored on a scale from 0 (least severe or impairing) to 4 (most severe or impairing). The 5-item distress scale was used as the primary outcome measure up to November 2016.

Hamilton Program for Schizophrenia Voices Questionnaire (*HPSVQ*) (Van Lieshout *et al.*, 2007). The HPSVQ covers similar topics to the PSYRATS-AH but in a self-report format over nine items. The physical and negative impact scales have been confirmed through factor analysis (Kim *et al.*, 2010). The four items of the negative impact scale were as follows: Q2, How bad are the things the voices say to you?; Q5, How much do the voices interfere with your daily activities?; Q6, How distressing are the voices that you hear?; and Q7, How bad (worthless/useless) do the voices make you feel about yourself? Each item was scored on a scale from 0 (least severe or impairing) to 4 (most severe or impairing). Kim *et al.* (2010) demonstrated high levels of reliability: Cronbach's alphas for the nine items were reported as 0.827, 0.913 and 0.940 at baseline, 1 week and 6 months; test–retest ICCs for the negative impact items ranged from 0.653 to 0.775 (good–excellent); HPSVQ scores also correlated highly ($r > 0.8$) with PSYRATS-AH scores. The 4-item negative impact scale was used as the primary outcome from November 2016.

Secondary outcomes

Choice of Outcome in CBT for Psychoses (*CHOICE*). The CHOICE short-form is a 12-item shortened version of Greenwood *et al.*'s (2010) self-report questionnaire assessing patient goals for CBT for psychosis that are relevant to subjective recovery. Items are rated on a 0–10 scale (0 = worst; 10 = best). The CHOICE short-form measure is currently being used nationally and internationally to evaluate outcomes following psychological therapies for psychosis, e.g. Jolley *et al.* (2015). The short form has recently been studied to determine its levels of reliability and validity. For copyright reasons, the exact psychometric properties cannot be published here ahead of the CHOICE short-form publication (Webb *et al.*, 2020; under review). However, the results demonstrated high internal consistency, high levels of sensitivity to change and good construct validity. This measure was included as a secondary outcome measure.

Depression Anxiety & Stress Scale (*DASS-21*) (Lovibond and Lovibond, 1995). This is a 21-item self-report instrument designed to measure the three negative emotional states of depression, anxiety and stress. Items are rated on a 4-point Likert scale of patients' experiences over the last week from 0 ('did not apply to me at all') to 3 ('applied to me very much, or most of the time'). Internal consistency and concurrent validity of the DASS-21 are reported to be in the acceptable to excellent ranges (Antony *et al.*, 1998). The DASS-21 was used up to November 2016.

Patient Health Questionnaire (*PHQ-9*). This is a 9-item self-report measure of depression symptom severity. Items are rated on a 4-point scale. Scores under 10 are considered sub-clinical, 10–14 mild, 15–19 moderate and 20+ severe. The scale has good levels of sensitivity (0.92) and specificity (0.78) (Gilbody *et al.*, 2007). The PHQ-9 was used from November 2016.

Procedure

PBCT was delivered as a 12-session group therapy, each session lasting 1.5 hours and delivered by two SVC therapists, one of whom was required by the local NHS trust clinical governance arrangements to be trained in the delivery of mindfulness-based interventions (13 therapists in total). The therapy manual is detailed elsewhere (Chadwick, 2006; Strauss and Hayward, 2013).

Each session began with a mindfulness practice and extended reflection. Mindfulness practice in PBCT is brief (10 minutes), with continuous guidance that includes reference to voice hearing experiences, and combines focused attention on body and breath with open awareness. Sessions 1–3 socratically drew out patients' voice hearing experiences (onset, impact, meaning, distress and coping) and framed them using the ABC cognitive model. Sessions 4–6 explored personal control, socratically identifying evidence from patients' experience that was not consistent with beliefs about voice power and control. Sessions 7–10 focused on identifying and decentring from negative schemata, and building positive schematic beliefs (including using experiential two-chair work) alongside recognition that the self is complex and changing. Sessions 11 and 12 explored learning and ways of taking this learning forward into everyday life. Patients were encouraged to practise mindfulness daily at home, using a supplied 5 or 10 minute recording.

Statistical analysis

Patient characteristics and clinical measures

Descriptive statistics, count (n), percentage (%), mean, standard deviation (SD), minimum and maximum were used to summarize patient characteristics and clinical measures pre- and post-PBCT group therapy, as appropriate. A single *voice-related distress* score was created by transforming the PSYRATS-AH distress scale and the HPSVQ negative impact scale into Z-scores and combining them into one variable; post-PBCT Z-scores were created using the corresponding baseline means and standard deviations. Similarly, a single *depression severity* score was created using Z-scores for PHQ-9 and the DASS-21 depression scale. We were able to combine the measures in this way as we were sufficiently confident that the paired scales measured the same constructs and were equally effective in detecting change. By combining the data we were potentially able to increase the statistical power of our exploratory analyses into the relationships within the data. The third clinical measure, *subjective recovery* (*CHOICE*), did not require any transformation.

Levels of engagement and predictors

Three levels of engagement were compared: (1) patients who were offered but did not commence therapy (non-commencers; attended 0 sessions); (2) patients who commenced but did not complete therapy (non-completers; attended 1–7 sessions); and (3) patients who completed therapy (completers; 8–12 sessions) (Chadwick *et al.*, 2016). To explore whether any of the patient characteristics or pre-PBCT clinical outcomes predicted level of engagement, one-way analysis of variance (ANOVA) and chi-square tests were used for continuous and categorical data, respectively. Fisher's exact tests were used where cross-tabulation cell counts were very small i.e. $n < 5$. Pre-PBCT scores for completers and non-completers were compared using a two-sample t-test. To improve the robustness of the statistical tests using categorical data, the following variables were re-categorized into binary variables: employment status (employed/not employed), relationship status (single/in a relationship), ethnic group (Black Asian and Minority Ethnic/White British) and diagnosis (psychosis/non-psychosis).

Clinical outcomes and predictors of change

A comparison of the pre- and post-PBCT clinical measures was carried out on completers using paired sample t-tests. The effect size was standardized using the t_c formula for Cohen's d (Dunlop *et al.*, 1996; Lenhard and Lenhard, 2016). In line with the methodology used in our previous work on predictors of change in completers (Paulik *et al.*, 2018), linear regression was used to assess whether any patient characteristics or pre-PBCT clinical measures predicted the change score after accounting for the baseline.

Results from all statistical tests were considered significant if $p < 0.05$ and 95% confidence intervals (CI) were created for all estimates. STATA version 16 was used for all analyses.

Reliable change and treatment response rate

Reliable change (RC; Jacobson and Traux, 1991) was calculated for each completer for each clinical measure (on their untransfomed scales) using baseline standard deviations (SD_{pre}) and corresponding Cronbach's alpha scores (pre-PBCT α_r) as follows: RC = (pre-PBCT score − post-PBCT score)/[$SD_{\mathrm{pre}}\sqrt{2}\sqrt{(1-\alpha_r)}$]. In accordance with the Jacobson–Traux methodology, RC scores larger than 1.96 are interpreted as real change above the fluctuations of an imprecise measuring instrument.

For each clinical measure, patients were categorized as either reaching a treatment response or not post-PBCT based on two cut-off levels. An improvement of at least 20% corresponded approximately to 'minimally improved', while a 50% reduction from baseline corresponded to 'much improved' (Bighelli *et al.*, 2018). Overall findings are summarized as treatment response rates.

Unplanned subgroup analyses

As an unplanned and exploratory analysis, patients with psychosis as a primary diagnosis or part of a mixed diagnosis (psychosis) were compared with those with no mention of psychosis in their diagnosis (non-psychosis) across all patient characteristics, clinical outcomes and levels of engagement. We felt that having an understanding of any systematic differences between the two subgroups would help with interpreting the findings from the comparison between the M4V (psychosis only) and the SVC (trans-diagnostic) populations. It would also help to highlight potential confounders.

Missing data

All missing data were assumed to be missing at random (Jakobsen *et al.*, 2017). Missing data were treated at the analysis stage using multiple imputation for chained equations (MICE; White *et al.*, 2011). The imputation model included all covariates used in the analysis model, covariates which predicted missingness and auxiliary variables that were correlated to the outcomes. The initial pool of auxiliary variables was age, duration of voice hearing, gender, education, employment status, ethnic group, relationship status and diagnosis. A sensitivity analysis was carried out by comparing the results of analyses using all available cases (complete case analysis) with those following multiple imputation (adjusted analysis).

Costs

The total cost of delivering a PBCT group was estimated based on two therapists (Agenda for Change Band 8a) and one clinic assistant (Agenda for Change Band 4) supporting each session with the following resources: clinic assistant administration (0.5 hours per session); therapist group delivery (3 hours per session), administration (preparation, liaison with care co-ordinators and writing notes; 2 hours per session) and clinical supervision (1 hour per month). The average cost per group was based on the total cost divided by the total number of groups delivered. The (2017) costs per hour of a clinic assistant and a therapist were £10.87 and £22.66, respectively.

Results

Patient characteristics

Ninety-five patients were invited to attend a PBCT group in SVC and were included in this study. Table 1 provides a descriptive summary of the characteristics of patients who were offered a place in a PBCT group. Fifty-two per cent ($n = 49$) of patients were female and 56% ($n = 50$) had a non-psychosis diagnosis, making this an atypical sample of patients who were distressed by voices. Other patient characteristics were distributed in line with previous studies of CBTp: the majority of patients were White British (75.0%; $n = 69$), had a mean age of 40.6 years ($SD = 11.9$; range = 19 to 67), did not have a partner (86.8%; $n = 73$), were unemployed (74%; $n = 66$), had been hearing voices for an average of 17.7 years ($SD = 14.8$; range = 0 to 61) and just under a third (31%; $n = 28$) of patients had left school at age 16 years or earlier.

Missing data

Counts of missing data for patient characteristics are given in the footnote for Table 1. For each clinical measure, data were missing in the following number of cases (n and %): voice hearing distress pre ($n = 6$; 6%) and post ($n = 39$; 41%); CHOICE pre ($n = 9$; 9%) and post ($n = 39$; 41%); depression severity pre ($n = 11$; 12%) and post ($n = 43$; 45%). Data were imputed for 164 missing data points (out of 1045; 16% missing) and the MICE model used 16 imputations. Predictors of missingness and auxiliary variables selected for the MICE model were gender, psychosis, education and duration of voices.

Engagement and predictors

The levels of engagement varied as follows: 59% ($n = 56$) completed therapy; 25% ($n = 24$) started but did not complete therapy with an attendance average of 3.9 sessions; 16% ($n = 15$) were offered but did not start the therapy. The attrition rate, therefore, was 41% ($n = 39$). Reasons for non-completer attrition were as follows (some gave multiple reasons): personal circumstances, e.g. moving house, changed job, divorce, family bereavement, finances ($n = 8$); health reasons/illness, e.g. surgery, anxiety ($n = 8$); did not like the group format ($n = 4$); symptoms were triggered during a session ($n = 4$); individual therapy was recommended by the therapist or requested by the patient ($n = 2$); feeling better ($n = 1$); no reason given ($n = 1$).

Statistical tests were applied to establish whether any patient characteristics or pre-PBCT clinical scores predicted level of engagement. Baseline depression severity Z-scores and voice-related distress Z-scores increased from lowest to highest for completers, non-completers and non-commencers, respectively. There was no such pattern for subjective recovery. A statistically significant between-group difference was only found between non-commencers and completers for the baseline depression severity Z-score (0.55; 95% CI –0.02 to 1.13, $SE = 0.29$; $p = 0.06$; $n = 65$) which changed to 0.59 (95% CI 0.01 to 1.18, $SE = 0.29$; $p = 0.047$; $n = 72$) after adjustment; non-commencers had higher levels of baseline depression. There were no other significant differences across the different groups.

Clinical outcomes and predictors of change in completers

A descriptive summary of all PBCT clinical measures (original and derived Z-scores) within each level of engagement is displayed in Table 2A (the breakdown for patient characteristics is not provided). Only summary statistics for completers are provided both pre- and post-PBCT as no data were collected post-PBCT for patients who did not commence or complete group therapy. Table 2B displays the Cronbach's alphas for each of the untransformed clinical measures to demonstrate the levels of internal consistency. Alpha ranges were 0.8029–0.9472 and 0.8852–0.9364 for pre-PBCT and post-PBCT measures, respectively, indicating good/excellent internal consistency.

Table 1. Descriptive summary of patient characteristics

Patient characteristic	Category	Count (%) $n = 95$
Age	Mean no. years (*SD*; range)	40.6 (11.9; 19–67)
Gender	Male	43 (45.74)
	Female	49 (52.13)
	Another term	2 (2.13)
	Total	94 (100)
Employment status		
	Employed FT/PT self	10 (11.24)
	Unemployed benefits	66 (74.16)
	Student	3 (3.37)
	Retired	4 (4.49)
	Home-maker	1 (1.12)
	Other	5 (5.62)
	Total	89 (100)
Relationship status		
	Single	60 (65.93)
	Cohabiting	5 (5.49)
	Widow	5 (5.49)
	Married/civil partner	11 (12.09)
	Separated/divorced	8 (8.79)
	Long-term relationship	2 (2.2)
	Total	91 (100)
Ethnic group		
	White British	69 (75)
	White other	9 (9.78)
	Asian	2 (2.17)
	Black	4 (4.35)
	Chinese	2 (2.17)
	Mixed	4 (4.35)
	Other	1 (1.09)
	Prefer not to say	1 (1.09)
	Total	92 (100)
Education		
	Left school <16	15 (16.67)
	Left School at 16	13 (14.44)
	Left School at 17/18	14 (15.56)
	College	30 (33.33)
	University	18 (20)
	Total	90 (100)
Age onset	Mean no. years (*SD*; range)	22.8 (13.0; 2–53)
Voice hearing duration	Mean no. years (*SD*; range)	17.7 (14.8; 0–61)
Diagnosis*	None	2 (2.17)
	Schizophrenia	28 (30.43)
	Schizoaffective	8 (8.7)
	BPD/EUPD	21 (22.83)
	PTSD	2 (2.17)
	Depression	5 (5.43)
	Mixed	20 (21.74)
	Other	6 (6.52)
	Total	92 (100)
Psychosis	No	50 (55.56)
	Yes	40 (44.44)
	Total	90 (100)

Missing data for age ($n = 3$), gender ($n = 1$), employment ($n = 6$), relationship status ($n = 4$), ethnicity ($n = 3$), education ($n = 5$), age at onset ($n = 6$), voice-hearing duration ($n = 6$) and diagnosis ($n = 3$). *Confirmed from psychiatrist's diagnosis notes.

Table 2A. Descriptive summary of PBCT clinical measures by level of engagement

		Pre-PBCT				Post-PBCT
		Completer	Non-completer	Non-commencer	Total	Completer
PSYRATS-AH Distress scale	*n*	31	15	10	56	26
	Mean	13.9	14.9	14.4	14.3	11.2
	SD	4.5	5.1	5.7	4.8	5.5
HPSVQ Negative impact scale	*n*	22	7	4	33	20
	Mean	11.3	11.7	14.5	11.8	8.7
	SD	3.6	2.8	1.3	3.4	4.6
Voice-related distress Z-score	*n*	53	22	14	89	46
	Mean	–0.1*	0.1	0.3	0	–0.8*
	SD	1.0	1.0	1.1	1.0	1.2
CHOICE (subjective recovery)	*n*	51	21	14	86	46
	Mean	4.18	3.98	4.22	4.14	4.98
	SD	2.0	2.0	1.8	2.0	1.9
DASS Depression	*n*	30	15	11	56	24
	Mean	11.5	13.5	15.4	12.8	9.6
	SD	6	6.8	6.3	6.3	5.1
PHQ-9	*n*	20	4	4	28	18
	Mean	17.9	19.2	20.5	18.5	16.8
	SD	6.0	5.9	4.5	5.7	6.8
Depression severity Z-score	*n*	50	19	15	84	42
	Mean	–0.2*	0.1	0.4	0	–0.4*
	SD	1.0	1.0	0.9	1.0	1.0

Statistics are count (*n*), mean and standard deviation (*SD*); *voice-related distress and depression severity have negative means because they are Z-scores.

Table 2B. Cronbach alphas for observed clinical measures, pre- and post-PBCT

Clinical measures	Pre-PBCT α_r	Post-PBCT α_r
PSYRATS-AH Distress scale	0.8590	0.9047
HPSVQ Negative impact scale	0.8029	0.8949
CHOICE (subjective recovery)	0.9472	0.9364
DASS Depression	0.9462	0.9003
PHQ-9	0.8267	0.8852

George and Mallery (2003; p. 231) provide the following rule of thumb: ≥0.9, excellent; ≥0.8, good; ≥0.7, acceptable; ≥0.6, questionable; ≥0.5, poor; ≤0.5, unacceptable.

Table 3 displays a summary of patient's observed change, reliable change and the treatment response rates. Overall, significant numbers of patients observed improvements in their clinical scores post-PBCT. For the primary outcome, reliable change was indicated for 29% of patients. However, on average, 51% and 29% of patients reached a 20% (minimally improved) and 50% (much improved) reduction in voice-related distress. Reliable change based on the depression score was indicated for 17% of patients. Overall, 28% of patients minimally improved and 8% were much improved. For subjective recovery, reliable change on the CHOICE was indicated for 28% of patients. However, 44% of patients minimally improved and 23% were much improved. Observed rates of deterioration on the RC scale were negligible at 2%, 3% and 2% for distress, depression and subjective recovery, respectively.

Pre–post differences in key outcomes for completers are displayed in Table 4. In the complete case analysis, PBCT had a medium-sized standardized effect on voice-related distress, which was reduced to an extent that was statistically significant (–0.61, 95% CI –0.95 to –0.28; $p < 0.001$; $d = -0.53$ (95% CI –0.88 to –0.23). After adjustment for missing data, the standardized effect

Table 3. Summary of pre–post observed change, reliable change and response rate in completers

	Total count	Improved *n* (%)				Unchanged *n* (%)		Deteriorated *n* (%)	
Clinical measure	*n*	Observed	RC	Minimally*	Much*	Observed	RC	Observed	RC
PSYRATS-AH Distress scale	25	18 (72.0)	6 (24.0)	12 (48.0)	7 (28.0)	1 (4)	18 (72.0)	6 (24)	1 (4.0)
HPSVQ Negative impact scale	16	11 (68.8)	6 (37.5)	9 (56.3)	5 (31.3)	4 (25.0)	10 (62.5)	1 (6.3)	0 (0)
Voice-related Distress (all)	41	29 (70.7)	12 (29.2)	21 (51.2)	12 (29.3)	5 (12.2)	28 (68.3)	7 (17.1)	1 (2.4)
CHOICE (subjective recovery)	43	25 (58.1)	12 (27.9)	19 (44.2)	10 (23.3)	1 (2.3)	30 (70)	17 (39.5)	1 (2.3)
DASS Depression	23	12 (52.2)	4 (17.4)	6 (26.1)	2 (8.7)	4 (17.4)	18 (78.3)	7 (30.4)	1 (4.3)
PHQ-9	13	7 (53.8)	2 (15.4)	4 (30.8)	1 (7.7)	3 (23.1)	11 (84.6)	3 (23.1)	0 (0)
Depression (all)	36	19 (52.8)	6 (16.7)	10 (28.0)	3 (8.3)	7 (19.4)	29 (80.5)	10 (27.8)	1 (2.8)

'Observed', observed pre–post change where improved means change > 0, unchanged means change = 0 and deterioration means change < 0; RC, reliable change where improved means RC > 1.96, unchanged means –1.96 < RC < 1.96 and deterioration means RC < –1.96; 'Minimally', minimally improved where there is a ≥20% change in pre–post observed change; 'Much', ≥50% change; *this method does not separate out those who are unchanged or who have deteriorated.

Table 4. Pre–post effect sizes in completers by key clinical outcomes

Clinical measure	Analysis type	Difference (95% CI)	*SE*	*t* (d.f.)	*p*-value	Cohen's *d* (95% CI)
Voice-related distress	Complete case	–0.61 (–0.95, –0.28)	0.17	–3.69 (43)	<0.001	–0.53 (–0.88, –0.23)
	Adjusted	–0.66 (–1.03, –0.29)	0.18	–3.63 (55)	0.001	–0.47 (–0.74, –0.20)
Depression severity	Complete case	–1.17 (–2.42, 0.07)	0.60	–2.18 (38)	0.0626	–0.20 (–0.42, –0.03)
	Adjusted	–0.24 (–0.48, –0.01)	0.12	–2.10 (55)	0.044	–0.20 (–0.39, –0.01)
Subjective recovery	Complete case	0.74 (0.30, 1.18)	0.22	3.41 (42)	0.0014	0.37 (0.14, 0.60)
	Adjusted	0.76 (0.31, 1.21)	0.22	3.45 (55)	0.001	0.35 (0.14, 0.56)

SE is standard error; *t* (d.f.) is the statistical test value and degrees of freedom; Cohen's *d* 95% CIs calculated based on non-central *t* distribution.

for the voice-related distress was slightly lower but remained statistically significant ($d = -0.47$, 95% CI –0.74 to –0.20; $p = 0.001$). The complete case analysis also revealed a small and statistically significant standardized effect on subjective recovery ($d = 0.37$, 95% CI 0.14 to 0.60, $p = 0.0014$). The small standardized effect on depression severity was non-significant. However, after adjustment for missing data, both depression severity and subjective recovery were statistically significant with small standardized effect sizes of $d = -0.2$ (95% CI –0.39 to –0.01; $p = 0.044$) and $d = 0.35$ (95% CI 0.14 to 0.56; $p = 0.001$), respectively.

Further analyses of predictors of change showed that among the completers, change in voice-related distress was predicted by pre-PBCT CHOICE and Depression severity. The regression coefficient for pre-PBCT CHOICE was –0.21 (95% CI –0.38 to –0.03; $SE = 0.09$; $p = 0.022$) which after adjustment decreased to –0.18 (95% CI –0.364 to –0.002; $SE = 0.09$, $p = 0.048$). The regression coefficient for pre-PBCT Depression severity was 0.41 (95% CI 0.01 to 0.81; $SE = 0.20$, $p = 0.046$) which after adjustment decreased to 0.34 (95% CI –0.04 to 0.72; $SE = 0.19$, $p = 0.079$), thus becoming non-significant. These findings indicate that patients with higher levels of subjective recovery or lower levels of depression before they started the PBCT groups experienced larger reductions in distress. None of the patient characteristics predicted change in voice-related distress, depression severity or subjective recovery.

Unplanned subgroup analyses

As an exploratory analysis, the 50 (56%) patients with a non-psychosis diagnosis were compared with the 40 (44%) with a psychosis diagnosis in terms of patient characteristics, engagement and clinical outcomes. There were no statistically significant differences between the two groups on patient characteristics or levels of engagement. The only observed statistically significant differences were on clinical scores at baseline: the psychosis group had smaller depression Z-scores (–0.781, 95% CI –1.19 to –0.37; $SE = 0.207$; $t = -3.77$; $p < 0.001$) and higher subjective recovery rates (1.30, 95% CI 0.48 to 2.13; $SE = 0.413$; $t = 3.15$; $p = 0.002$). However, there were no baseline differences for distress scores. Statistical tests for between-group differences on the levels of change in outcome, after accounting for the baseline scores, were non-significant for all clinical measures.

Sensitivity analysis

Across all analyses, results obtained using the observed data have been compared with those from the imputed data. Only on two occasions did conclusions differ, but this was due to the strict adherence to the 5% alpha cut-off which in exploratory analyses such as these is arguably conservative (Fisher, 1950). The significance level for the pre–post Depression severity score difference decreased from $p = 0.0626$ (complete case) to $p = 0.044$ (adjusted). In the predictor of change in voice-related distress analysis, the significance for the pre-PBCT Depression severity coefficient increased from $p = 0.046$ (complete case) to $p = 0.079$ (adjusted).

Resources

The total cost for each 12-session PBCT group was estimated to be £1288, which equates to £130 per session and the overall cost per patient was £214.

Discussion

This study examined engagement, outcomes and costs of PBCT groups for distressing voices as part of a therapy pathway within a routine clinical setting. It also explored the predictive value of pre-therapy clinical measures and patient characteristics in determining the level of engagement and outcomes for patients who completed therapy. A total of 56% of patients had a non-psychosis diagnosis, which made this an atypical sample of patients who were distressed by voices.

The majority of patients, 59% (95% CI 48 to 69), completed therapy by attending at least eight of the 12 group sessions. We predicted that this completion rate would be lower than the 72% (95% CI 58 to 84) observed in the M4V trial. We hypothesized that the higher trial rate could be attributable to the availability of additional resources within a funded study, compared with routine clinical practice, and their use to maintain engagement (i.e. by providing transport to sessions). However, although the SVC rate was smaller, the difference was statistically non-significant at the 5% alpha level ($p = 0.1126$; 35% power). The level of engagement in the current study was associated with baseline levels of depression and voice-related distress (with lower levels indicating better engagement). The only significant difference between the engagement groupings was found for depression (with non-commencers reporting higher levels of depression than completers).

Patients who completed the group therapy experienced statistically significant medium-sized reductions in voice-related distress ($d = -0.47$, 95% CI –0.74 to –0.20) and, contrary to our hypothesis, this was comparable (in terms of the overlapping confidence intervals) to the medium-sized reductions observed in the M4V trial ($d = -0.46$, 95% CI –0.93 to –0.01). These outcomes are on a par with the medium effect sizes generated by NICE-compliant CBTp offered within the clinical setting of the PICuP clinic in South London (Peters *et al.*, 2015). They also offer further evidence of the ability of therapists in routine clinical practice to generate outcomes that are similar to those generated in research environments. However, the atypical nature of these clinical environments should be acknowledged as SVC and PICuP offer a level of training, support and supervision that may not be available elsewhere. Furthermore, these outcomes can be viewed from varying perspectives, each of which may seem to tell a slightly different story as: the majority of patients reported an improvement on the primary outcome (71%); about half experienced a treatment response in the minimum improvement category (51%); and reliable change was demonstrated by only a minority (29%). These varying perspectives are reflective of the ongoing debate within the field about which outcomes to measure (Thomas *et al.*, 2014) and how they should be measured (Badcock *et al.*, 2020).

There were small but statistically significant improvements in subjective recovery $d = 0.35$ (95% CI 0.14 to 0.56) and depression severity $d = -0.20$ (95% CI –0.39 to –0.01). Equivalent treatment effect sizes in the M4V trial were larger for subjective recovery $d = 0.47$ (95% CI 0.001 to 0.937) and depression $d = -0.60$ (95% CI –1.08 to –0.13). In terms of predictors of change, greater reductions in voice-related distress at post-therapy were related to higher pre-therapy levels of subjective recovery. Reductions in voice-related distress were not linked to any of the other pre-therapy measures or patient characteristics.

Our findings for predictors of engagement and change suggest the following: patients with greater well-being at the start of the group are more able to engage and experience better outcomes. The two observed correlations (depression with engagement; subjective recovery with outcome) suggest that differing factors may impact at different points along the patient's therapeutic journey. High levels of depression may adversely affect a patient's motivation or willingness to commence the therapy. Meanwhile, low levels of subjective recovery may impede learning despite appropriate levels of attendance at sessions. In line with Paulik *et al.* (2018), high levels of depression and poor subjective recovery may indicate the need for some pre-therapy work to try to lift mood to facilitate recovery.

The trans-diagnostic nature of the SVC sample was noteworthy but the presence/absence of a psychosis diagnosis seemed to have little impact upon engagement and outcome. Planned comparisons across diagnoses within a larger study would be required to further explore any possible associations between diagnoses and engagement/outcome.

The evaluation of the resources required to deliver PBCT groups within SVC is suggestive of good value. The cost of £214 per patient is very small relative to the cost of £404 for one day spent in an inpatient psychiatric unit (Department of Health, 2017). Value is also apparent when the sessional cost of £130 is compared with the £133 cost of a session of individual CBTp (Sheaves *et al.*, 2019). In addition, PBCT offers 90-minute group sessions as well as the opportunity for learning from peers. The potential for CBT to save money when caring for psychosis patients has been reported by NICE (2014) and Sheaves *et al.* (2019) after accounting for the cost of therapy, and future research could explore any relative savings within SVC.

Limitations

Firstly, as this was a service evaluation there was no control group so we cannot, for instance, assess whether patients would have got better anyway regardless of therapy. We therefore cannot rule out the possibility that the pre–post changes in clinical measures were attributable to factors unrelated to the PBCT group. As with all uncontrolled pre–post studies, our findings may be misleading due to regression to the mean. Although it is unlikely, the pooling of our data to create Z-scores may have altered the variability in the data and led to spurious results. Also, as there was no randomization, there is a higher risk of bias in our results. This could include conscious or unconscious selection bias whereby patients were referred to SVC by their care co-ordinator based on their own preconceived ideas. For example, a care co-ordinator may be more cautious about referring patients they deemed as too unwell even though they would have met the SVC eligibility criteria. As a result, SVC may have been more likely to receive referrals of patients predicted to benefit. Our findings around outcomes in particular may therefore be over-estimated. Although encouraging findings were found in terms of improved outcomes, we have no indication of the extent to which these outcomes would be maintained at follow-up. Finally, this study is service and context specific and so the ability for generalization is limited.

Clinical implications

This study demonstrates that evidence-based psychological therapy groups can be delivered trans-diagnostically in routine clinical practice to patients distressed by hearing voices, with the amount of benefit comparable to that generated in a research environment. In order to optimize the use of resources, we recommend that attention should be paid to baseline levels of depression and subjective recovery, as they have the potential to influence engagement and outcomes, respectively. The costs for PBCT groups are relatively low and so this therapy offers a potentially cost-effective intervention for a group of patients who may otherwise make extensive use of expensive inpatient facilities. Future research should include a longitudinal study to establish the level and duration of maintained gains with an embedded subgroup analyses to better understand what works for whom.

Acknowledgements. We would like to express our thanks to the many therapists and clinic assistants within the Sussex Voices Clinic who have delivered therapy and collected data for this study.

Financial support. None.

Conflicts of interest. None.

Ethical statements. The authors have abided by the Ethical Principals of Psychologists and Code of Conduct as set out by the BABCP and BPS. In accordance with the UK Policy Framework for Health and Social Care Research (2017), NHS Research Ethics Committee approval was not required. The study was registered with an NHS Audit team in 2015.

References

Antony, M. M., Bieling, P. J., Cox, B. J., Enns, M. W., & Swinson, R. P. (1998). Psychometric properties of the 42-item and 21-item versions of the Depression Anxiety Stress Scales in clinical groups and a community sample. *Psychological Assessment*, *10*, 176–181.

Badcock, J. C., Graham, M. E., & Paulik, G. (2020). Assessing the clinical significance of treatment outcomes for distressing voices in routine clinical practice. *Clinical Psychology and Psychotherapy*, *27*, 79–86. https://doi.org/10.1002/cpp.2410

Bighelli, I., Huhn, M., Schneider-Thoma, J. ***et al.*** (2018). Response rates in patients with schizophrenia and positive symptoms receiving cognitive behavioural therapy: a systematic review and single-group meta-analysis. *BMC Psychiatry*, *18*, 380. https://doi.org/10.1186/s12888-018-1964-8

Chadwick, P. (2006). *Person-Based Cognitive Therapy for Distressing Psychosis*. Chichester, UK: John Wiley & Sons Ltd.

Chadwick, P., Strauss, C., Jones, A.-M., Kingdon, D., Ellet, L., Dannahy, L., & Hayward, M. (2016). Group mindfulness-based intervention for distressing voices: a pragmatic randomized controlled trial. *Schizophrenia Research*, *175*, 168–173.

Department of Health (2017). *UK Policy Framework for Health and Social Care Research*. London, UK: Department of Health.

Dunlop, W. P., Cortina, J. M., Vaslow, J. B., & Burke, M. J. (1996). Meta-analysis of experiments with matched groups or repeated measures designs. *Psychological Methods*, *1*, 170–177.

Fisher, R. A. (1950). *Statistical Methods for Research Workers*. London, UK: Oliver and Boyd.

George, D., & Mallery, P. (2003). *SPSS for Windows Step by Step: A Simple Guide and Reference. 11.0 update* (4th edn). Boston, USA: Allyn & Bacon.

Gilbody, S., Richards, D., & Barkham, M. (2007) Diagnosing depression in primary care using self-completed instruments: UK validation of PHQ-9 and CORE-OM. *British Journal of General Practice*, *57*, 650–652.

Goodliffe, L., Hayward, M., Brown, D., Turton, W., & Dannahy, L. (2010). Group person-based cognitive therapy for distressing voices: views from the hearers. *Psychotherapy Research*, *20*, 447–461.

Greenwood, K. E., Sweeney, A., Williams, S., Garety, P., Kuipers, E., Scott, J., & Peters, E. (2010). Choice of outcome in CBT for psychoses (CHOICE): the development of a new service user-led outcome measure of CBT for psychosis. *Schizophrenia Bulletin*, *36*, 126–135.

Haddock, G., McCarron, J., Tarrier, N., & Faragher, E. B. (1999). Scales to measure dimensions of hallucinations and delusions: the psychotic symptoms rating scales (PSYRATS). *Psychological Medicine*, *29*, 879–889.

Hayward, M., Edgecumbe, R., Jones, A.-M., Berry, C., & Strauss, C. (2018). Brief coping strategy enhancement for distressing voices: an evaluation in routine clinical practice. *Behavioural and Cognitive Psychotherapy*, *46*, 226–237.

Hayward, M., Jones, A.-M., Bogen-Johnston, L., Thomas, N., & Strauss, C. (2017). Relating therapy for distressing auditory hallucinations: a pilot randomized controlled trial. *Schizophrenia Research*, *183*, 137–142.

Hazell, C. M., Greenwood, K., Fielding-Smith, S., Rammou, A., Bogen-Johnston, L., Berry, C., Jones, A. M., & Hayward, M. (2018a). Understanding the barriers to accessing symptom-specific cognitive behavior therapy (CBT) for distressing voices: reflecting on and extending the lessons learnt from the CBT for psychosis literature. *Frontiers in Psychology*, *9*, 727. doi: 10.3389/fpsyg.2018.00727

Hazell, C. M., Hayward, M., Cavanagh, K., Jones, A.-M., & Strauss, C. (2018b). Guided self-help cognitive-behaviour Intervention for VoicEs (GiVE): results from a pilot randomised controlled trial in a transdiagnostic sample. *Schizophrenia Research*, *195*, 441–447.

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.

Jakobsen, J. C., Gluud, C., Wetterslev, J., & Winkel, P. (2017). When and how should multiple imputation be used for handling missing data in randomised clinical trials – a practical guide with flowcharts. *BMC Medical Research Methodology*, *17*, 162.

Jolley, S., Garety, P., Peters, E., Fornells-Ambrojo, M., Onwumere, J., Harris, V., . . . & Johns, L. (2015). Opportunities and challenges in improving access to psychological therapies for people with severe mental illness (IAPT-SMI): evaluating the first operational year of the South London and Maudsley (SLaM) demonstration site for psychosis. *Behavioural Research and Therapy*, *64*, 24–30.

Kay, S. R., Fiszbein, A., & Opler, L. A. (1987). The positive and negative syndrome scale for schizophrenia. *Schizophrenia Bulletin*, *13*, 261–276.

Kim, S. H., Jung, H. Y., Hwang, S. S., Chang, J. S., Kim, Y., Ahn, Y. M., & Kim, Y. S. (2010). The usefulness of a self-report questionnaire measuring auditory verbal hallucinations. *Progress in Neuro-Psychopharmacology & Biological Psychiatry*, *34*, 968–973.

Lenhard, W., & Lenhard, A. (2016). Calculation of effect sizes. Retrieved from: https://psychometrica.de/effect_size.html. Dettelbach (Germany): Psychometrica. doi: 10.13140/RG.2.1.3478.4245

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

McHale, C., Hayward, M., & Jones, F. W. (2018). Building a grounded theory of engagement in mindfulness-based group therapy for distressing voices. *Qualitative Health Research*, *28*, 2169–2182.

Morera, T., Bucci, S., Randal, C., Barrett, M. & Pratt, D. (2015). Exploring views about mindfulness groups for voice-hearing from the perspective of service users and staff: a Q-methodology study. *Psychotherapy Research*, *27*, 179–188.

National Institute for Health and Care Excellence (NICE) (2014). *Psychosis and Schizophrenia in Adults: Treatment and Management* (CG178). London, UK: National Institute for Health and Care Excellence. Available at: http://guidance.nice.org.uk/CG178

Payne, T., Allen, J., & Lavender, T. (2017). Hearing Voices Network groups: experiences of eight voice hearers and the connection to group processes and recovery. *Psychosis*, *9*, 205–215.

Paulik, G., Jones, A.-M., & Hayward, M. (2018). Brief coping strategy enhancement for distressing voices: predictors of engagement and outcome in routine clinical practice. *Clinical Psychology and Psychotherapy*, *25*, 634–640.

Peters, E., Crombie, T., Agbedjro, D., Johns, L. C., Stahl, D., Greenwood, K., … & Kuipers, E. (2015). The long-term effectiveness of cognitive behavior therapy for psychosis within a routine psychological therapies service. *Frontiers in Psychology*, *6*, 1658.

Royal College of Psychiatrists (2018). *National Clinical Audit of Psychosis – National Report for the Core Audit.* London, UK: Healthcare Quality Improvement Partnership. https://www.hqip.org.uk/wp-content/uploads/2018/07/NCAP-National-report-for-core-audit-2018-FINAL.pdf

Sheaves, B., Peters, E., Stahl, D., & Johns, L. (2019). Changes in care costs associated with cognitive behavioural therapy for psychosis delivered in routine clinical practice. *Journal of Mental Health.* doi: 10.1080/09638237.2019.1581327

Strauss, C., & Hayward, M. (2013). Group person-based cognitive therapy for distressing psychosis. In L. Johns, E. Morris & J. Oliver (eds), *ACT and Mindfulness for Psychosis.* London, UK: Wiley-Blackwell.

Thomas, N., Hayward, M., Peters, E., van der Gaag, M., Bentall, R.P., Jenner, J. *et al.* (2014). Psychological therapies for auditory hallucinations (voices): current status and key directions for future research. *Schizophrenia Bulletin*, *40* (suppl. 4), S202–S212.

Van Lieshout, R. J., & Goldberg, J. O. (2007). Quantifying self-reports of auditory verbal hallucinations in persons with psychosis. *Canandian Journal of Behioural Science*, *39*, 73–77.

Webb, R., James, B., Skan, R., Peters, E., Garety, P., Kuipers, E., Hayward, M., & Greenwood, K. (2020) Exploring the development, validity and utility of the Short Form version of the Choice of Outcome In CBT for psychosEs (CHOICE-SF): a patient reported outcome measure of psychological recovery (under review).

White, I. R., Royston, P., & Wood, A. M. (2011). Multiple imputation using chained equations: issues and guidance for practice. *Statistics in Medicine*, *30*, 377–399.

Woodward, T. S., Jung, K., Hwang, H., Yinm J., Taylor, L., Menon, M., … & Erickson, D. (2014). Symptom dimensions of the psychotic symptom rating scales in psychosis: a multisite study. *Schizophrenia Bulletin*, *40* (suppl. 4), 265–274.

Cite this article: Jones A-M, Strauss C, and Hayward M (2021). A service evaluation of a group mindfulness-based intervention for distressing voices: how do findings from a randomized controlled trial compare with routine clinical practice? *Behavioural and Cognitive Psychotherapy* **49**, 76–90. https://doi.org/10.1017/S1352465820000624