Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Smartphone Cognitive Bias Modification for Paranoia: A Randomized Controlled Trial of STOP

Jenny Yiend

jenny.yiend@kcl.ac.uk

King's College London https://orcid.org/0000-0002-1967-6292

Rayan Taher

King's College London https://orcid.org/0000-0001-8592-527X

Carolina Fiahlo

King's College London

Chloe Hampshire

University of Bath

John Hodsoll

King's College London

Che Hsu

Department of Psychological Medicine, Dunedin School of Medicine, University of Otago, New Zealand.

Thomas Kabir

University of Oxford

Jeroen Keppens

King's College London

Philip McGuire

University of Oxford

Emmanuelle Peters

King's College London

Tanya Ricci

King's College London

Sukhwinder Shergill

Kent and Medway Medical School

Daniel Stahl

King's College London https://orcid.org/0000-0001-7987-6619

George Vamvakas

King's College London

MPIT Group

McPin Public Involvement Team

Avegen Group

Avegen Health

Pamela Jacobsen

University of Bath

Article

Keywords: Cognitive Bias Modification; Transdiagnostic Paranoia; Digital Mental Health Interventions; Randomized Controlled Trial; Paranoia Scale

Posted Date: November 19th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11748854/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License

Additional Declarations: Yes there is potential Competing Interest. The digital intervention evaluated in this study was developed by Avegen Health Ltd., who received funding for its development from King's College London. The authors of this paper are independent from both the developers and the funders.

Smartphone Cognitive Bias Modification for Paranoia: A Randomized Controlled Trial of STOP

Jenny Yiend[1*#], Rayan Taher[1*], Carolina Fialho[1], Chloe Hampshire[2], John Hodsoll[9], Che-Wei Hsu[1,3,4], Thomas Kabir[5], Jeroen Keppens[6], Philip McGuire[5], Emmanuelle Peters[1,7], Tanya Ricci[1], Sukhwinder Shergill[8], Daniel Stahl[9], George Vamvakas[9], the MPIT[10], Avegen[11], Pamela Jacobsen[2].

[1] *Institute of Psychiatry, Psychology & Neuroscience, King's College London, London, UK*
[2] *Department of Psychology, University of Bath, Bath, UK.*
[3] *Department of Psychological Medicine, Dunedin School of Medicine, University of Otago, NZ.*
[4] *College of Community Development and Personal Wellbeing, Otago Polytechnic, NZ.*
[5] *Department of Psychiatry, University of Oxford, Oxford, UK.*
[6] *Department of Informatics, King's College London, London, UK.*
[7] *South London and Maudsley NHS Foundation Trust, London, UK*.
[8] *Kent and Medway Medical School, Canterbury, Kent, UK.*
[9] *Department of Biostatistics and Health Informatics*, *Institute of Psychiatry, Psychology & Neuroscience, King's College London, UK*.
[10] *McPin Public Involvement Team includes, in alphabetical order: Al Richards, Alexandra Kenny, Edmund Brooks, Emily Curtis, and Vanessa Pinfold.*
[11] *The Avegen group includes, in alphabetical order: Nandita Kurup, Neeraj Apte,*

*Joint first authors

Corresponding author

Keywords: Cognitive Bias Modification; Transdiagnostic Paranoia; Digital Mental Health Interventions; Randomized Controlled Trial; Paranoia Scale

Abstract

Paranoia, distressing beliefs about harm from others, crosses diagnostic boundaries and impairs daily functioning. STOP is a self-administered smartphone app delivering cognitive bias modification therapy targeting paranoia. In a double-blind randomized controlled trial, with 21 national recruiting sites, 274 participants with clinically significant paranoia were given 6- or 12-weeks of STOP. Adherence was high (85%), with no intervention-related serious adverse events. Both doses were effective in reducing paranoid bias (ds = -.29 to -.62). The 12-week dose showed the strongest symptom effects, with significant reductions in paranoia on the primary outcome compared to control post treatment ($d = -0.33$, $p = 0.026$), although the primary endpoint (24 weeks) was not significant ($d = -0.24$, $p = 0.141$). Significant improvements were observed at 24 weeks in paranoia-related worry ($d = -0.45$, $p = 0.001$), with trend level gains in anxiety, depression and recovery. These findings suggest that self-guided digital cognitive bias modification is safe and acceptable, supporting the possibility that an improved version of STOP may provide benefits in the treatment of paranoia. Trial registration: ISRCTN17754650.

Introduction

Digital mental health interventions offer a scalable route to address persistent treatment gaps in mental health care, including for serious mental illness[1–3]. Although evidence remains heterogeneous, systematic reviews and meta-analyses indicate that several approaches are feasible, acceptable and show preliminary efficacy for psychosis-spectrum symptoms[3], with randomized trials reporting clinically meaningful reductions in voice-related distress, persecutory delusions, belief flexibility and symptom-management outcomes[4–9]. However, most digital interventions have targeted broad psychosis-spectrum presentations or specific symptom domains, and few have been designed to reduce paranoia as a primary, transdiagnostic outcome. There is therefore a need for rigorously evaluated digital interventions tailored to paranoia across diagnostic boundaries.

Paranoia is the belief that others intend harm, including emotional harm, such as humiliation, or physical harm, such as a threat to one's safety. Psychological models conceptualize paranoid thinking as a continuum present in both the general population and clinical groups, with a hierarchical distribution from relatively common social-evaluative concerns, such as fears of rejection, to less common beliefs of severe threat[10–14]. Paranoia occurs across multiple forms of psychopathology, including depression[15], bipolar disorder[16], post-traumatic stress disorder[17] and anxiety[18], but is most prevalent and severe in psychosis[19]. Persecutory delusions represent the extreme end of this continuum and are associated with greater distress than other delusion types[20], increased likelihood of harmful action towards self or others[21], and hospitalization[22]. More than one-third of UK psychiatric patients experience persecutory delusions[23]. Although UK guidelines recommend cognitive behavioural therapy, access remains limited, with uptake reported in as few as 24% of those who could benefit[24–26],

and many continue to experience distressing symptoms after psychological treatment[27,28]. This has sharpened interest in briefer, mechanism-focused interventions targeting putative causal factors such as cognitive biases[29,30].

Here, bias refers to selective processing of information that reflects the core content of a disorder[14]. Cognitive Bias Modification for Interpretation is a theory-driven approach that uses repeated word-based tasks to shift interpretation towards more adaptive meanings and has shown benefit in anxiety and depression[31]. In paranoia, interpretation bias may lead ambiguous social events to be perceived as personally threatening, for example, hearing people laugh nearby may be taken as ridicule rather than coincidence. CBM-pa, a six-session desktop intervention targeting paranoid interpretations, showed promising feasibility[32,33]. STOP (Successful Treatment of Paranoia) extends this approach to a 12-session, self-administered mobile app with audio and graphic enhancements[34] (see Methods, Figure 5).

Here we report the main outcomes of the STOP randomized controlled trial, which evaluated the efficacy and safety of a self-administered intervention for transdiagnostic use in people experiencing distressing clinical levels of paranoia. Efficacy was tested by comparing a 6-week dose of STOP (STOP 6), and a 12-week dose of STOP (STOP 12), with a 12-week control (CTRL 12). CTRL 12 was a custom built, text-reading app matched for app usage, research team interactions and other nonspecific placebo effects of trial participation. The primary outcome was the Paranoia Scale (PS) score at 24 weeks post randomization[35]. Safety was assessed using the Negative Effects Questionnaire at 12 weeks and fortnightly adverse-event checklists throughout the intervention period[36].

Results

Participants

Between 1st October 2021 and 15th June 2024 we received 1,017 referrals to the trial from secondary care (953; 77.8%), the third sector (116; 9.5%) and by self-referral through the study website (156; 12.7%). Of these 452 were assessed for eligibility. 569 did not proceed from referral to eligibility assessment for the following reasons: 225 lost interest, 239 became uncontactable, 62 became ineligible, and 43 were lost for other reasons. 274 participants met the eligibility criteria and were randomly assigned to receive either 6 weeks of STOP (n= 92), 12 weeks of STOP (n= 90), or a placebo control (n=92) (see Figure 1 for the Consolidated Standards of Reporting Trials (CONSORT) diagram). Data were available for 259 (95%) participants at the final 24 week follow up: 8 participants were lost from the control group; 1 from the 6-week intervention and 6 from the 12-week intervention.

010 Flow Diagram

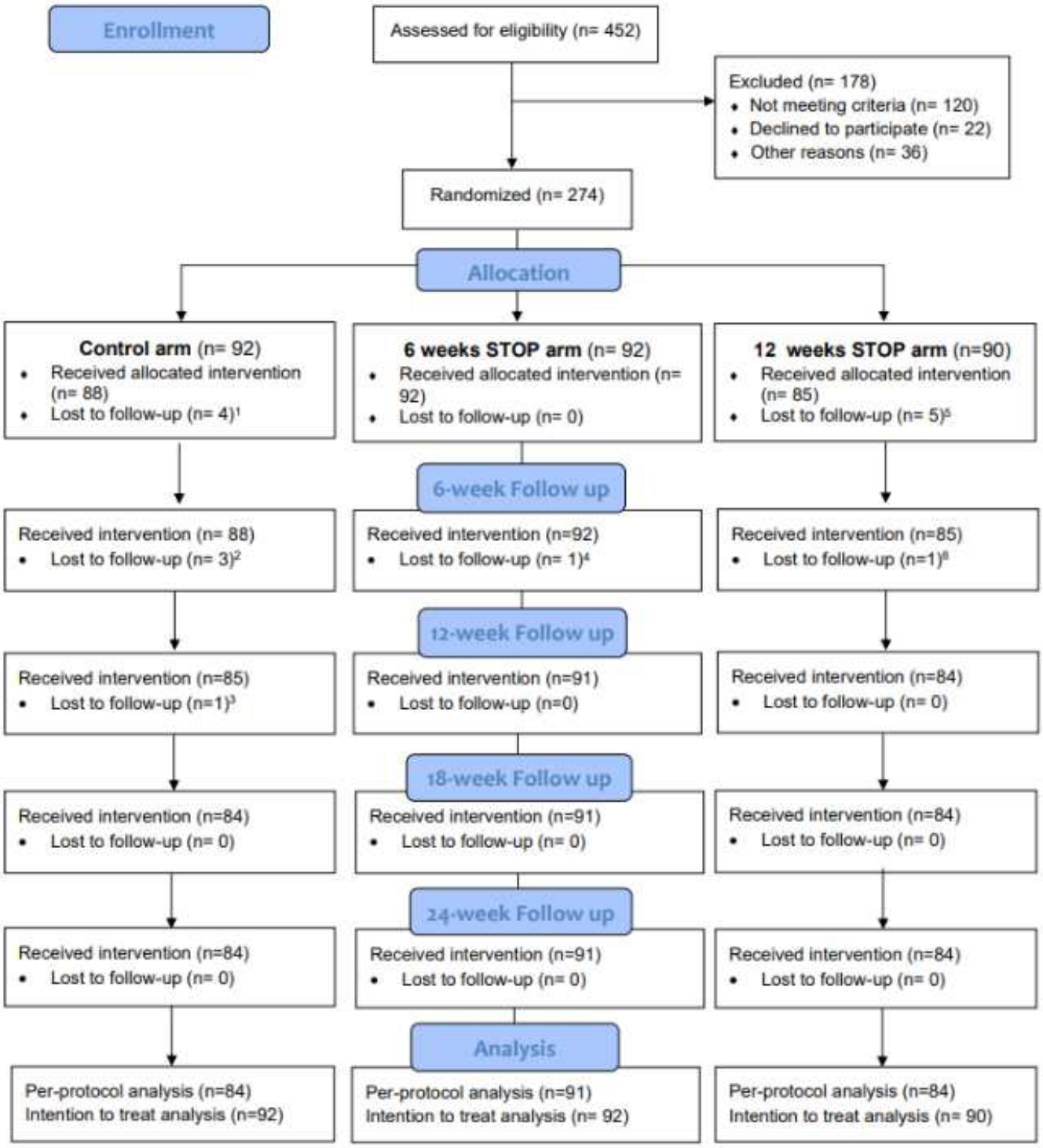


Fig. 1 CONSORT Diagram. Available cases include all participants with at least one post-randomisation outcome.1. Deterioration in mental health (n=2) and worsened medical condition (n=2). 2. Practical Burden (n=2) and Withdrawn consent (n=1). 3. Practical burden (n=1). 4. Worsened medical condition (n=1). 5. Withdrawn consent (n=3) and Practical burden (n=2). 6. Deterioration in mental health (n=1). *Lost to Follow up is defined as withdrawals in this diagram; occasional data missingness is reported elsewhere.

Participants' baseline demographic and clinical characteristics were mostly comparable between arms at baseline (Table 1). Across all arms slightly over half the sample were female (53%), single (63%), white (74.5%), unemployed (56%), and held a university degree (39%). This profile reflected the intended transdiagnostic reach and broad referral base of our target population. The average age was 40.8 years, with the average age of paranoia onset at 22.8 years. Participants in the control arm had a 3 to 4 years lower age of onset than the intervention arms.

Table 1. Baseline characteristics of trial participants overall and by intervention arm.

Baseline Characteristic	**All Arms** N=274	**STOP 6** N = 92	**STOP 12** N = 90	**CTRL 12** N = 92
Age				
Mean (SD)	40.8 (12.3)	40.4 (13.1)	41.1 (11.9)	40.9 (12.1)
Median (Min–Max)	41.0 (18.0–72.0)	40.0 (18.0–70.0)	42.0 (20.0–72.0)	40.5 (18.0–69.0)
Age of onset (Years)				
Mean (SD)	22.8 (9.6)	23.3 (10.0)	24.5 (10.7)	20.6 (7.5)
Median (Min–Max)	20.0 (13.0–64.0)	21.0 (13.0–64.0)	21.0 (13.0–58.0)	19.0 (13.0–45.0)
Gender				
Male	128 (47%)	42 (46%)	40 (44%)	46 (50%)
Female	145 (53%)	49 (53%)	50 (56%)	46 (50%)
Other	1 (0.4%)	1 (1.1%)	0 (0%)	0 (0%)
Ethnicity				
White	203 (74.1%)	71 (25.9%)	69 (25.2%)	63 (23%)
Mixed	21 (7.7%)	6 (2.2%)	4 (1.5%)	11 (4.0%)
Asian	16 (5.8%)	6 (2.2%)	4 (1.5%)	6 (2.2%)

Black,	29 (10.6%)	7 (2.6%)	11 (4.0%)	11 (4.0%)
Other	5 (1.8%)	2 (0.7%)	2 (0.7%)	1 (0.4%)
Education Level				
No qualifications	18 (6.6%)	5 (5.4%)	7 (7.8%)	6 (6.5%)
GCSE/O-Levels	51 (19%)	14 (15%)	21 (23%)	16 (17%)
A-Levels	48 (18%)	17 (18%)	13 (14%)	18 (20%)
Vocational/College	51 (19%)	18 (20%)	19 (21%)	14 (15%)
University/Professional Qual (degree+)	106 (39%)	38 (41%)	30 (33%)	38 (41%)
Occupation				
Unemployed	153 (56%)	49 (53%)	51 (57%)	53 (58%)
Student	28 (10%)	9 (9.8%)	10 (11%)	9 (9.8%)
Full-time	48 (18%)	20 (22%)	13 (14%)	15 (16%)
Part-time	45 (16%)	14 (15%)	16 (18%)	15 (16%)
Dyslexia	37 (14%)	7 (7.6%)	17 (19%)	13 (14%)
Relationship Status				
Single	172 (63%)	61 (66%)	53 (59%)	58 (63%)
Married/Cohabiting	55 (20%)	16 (17%)	22 (24%)	17 (18%)
In a Steady Relationship	34 (12%)	11 (12%)	12 (13%)	11 (12%)
Divorced/Separated	10 (3.6%)	3 (3.3%)	2 (2.2%)	5 (5.4%)
Widowed	3 (1.1%)	1 (1.1%)	1 (1.1%)	1 (1.1%)
Living Arrangements				
Alone	121 (44%)	41 (45%)	39 (43%)	41 (45%)
Alone with children	14 (5.1%)	3 (3.3%)	3 (3.3%)	8 (8.7%)
Partner/Spouse	32 (12%)	10 (11%)	12 (13%)	10 (11%)
Partner/Spouse and Children	25 (9.1%)	6 (6.5%)	11 (12%)	8 (8.7%)
Parents	40 (15%)	15 (16%)	12 (13%)	13 (14%)
Other family	12 (4.4%)	5 (5.4%)	6 (6.7%)	1 (1.1%)
Friends	6 (2.2%)	2 (2.2%)	3 (3.3%)	1 (1.1%)

Other	24 (8.8%)	10 (11%)	4 (4.4%)	10 (11%)

Primary Outcome

As shown in Table 2 and Figs. 2 and 3, all three arms showed a decrease in paranoia symptoms on the PS after treatment. Results are reported as standardized effect sizes (ES; Cohen's d). At 12 weeks, both STOP 6 and STOP 12 showed a reduction in paranoia symptoms compared to CTRL 12, with STOP 12 demonstrating a slightly stronger effect, a (5.29 point drop , (ES = -0.33, 95% CI –0.64, -0.04, p = 0.026). By 24 weeks, the treatment effects for STOP 12 remained larger than for STOP 6 at 3.80 points (ES = -0.33, 95% CI –0.64, -0.04, p = 0.026) vs 2.3 points (ES = -0.14, 95% CI –0.44, 0.15). These differences were not statistically significant (p's > 0.14).

Table 2. Outcomes changes between baseline, 6, 12, 18, and 24 weeks: ITT analysis (N = 274). Adjusted group differences are linear contrasts from a linear mixed model for repeated measures, including fixed effects of site and gender. Standardised effect sizes (ES; Cohen's d) derived from dividing the adjusted difference by the baseline standard deviation of the outcome.

	Unadjusted Means							**Adjusted Group Differences**			
	STOP 06		**STOP 12**		**CTRL 12**			**STOP 6 vs CTRL 12**		**STOP 12 vs CTRL 12**	
	N	Mean (SD)	N	Mean (SD)	N	Mean (SD)	Joint Test $X^2(2)$ p- value	Adjusted Difference [95% CI] p-value	ES [95% CI]	Adjusted Difference [95% CI] p-value	ES [95% CI)
PS											
Baseline	92	67.8 (15.2)	90	66.2 (15.2)	92	65.4 (17.5)					
Week 6	80	58.2 (21.0)	73	54.8 (20.4)	73	58.3 (21.2)	$X^2(2) = 4.91$ p = 0.086	-2.55 [-6.92, 1.83] p = 0.254	-0.16 [-0.43, 0.11]	-4.64 [-8.75, -0.53] **p = 0.027**	-0.29 [-0.55, -0.03]
Week 12	83	53.1 (19.2)	66	52.6 (19.7)	69	55.5 (21.5)	$X^2(2) = 6.12$ **p = 0.047**	-4.76 [-9.36, -0.16] **p = 0.043**	-0.30 [-0.59, -0.01]	-5.29 [-9.95, -0.63] **p = 0.026**	-0.33 [-0.62, -0.04]
Week 18	80	54.7 (19.9)	72	52.3 (20.4)	71	55.7 (22.9)	$X^2(2) = 2.99$ p = 0.224	-2.52 [-7.34, 2.30] p = 0.305	-0.16 [-0.46, 0.14]	-4.16 [-8.92, 0.60] p = 0.086	-0.26 [-0.56, 0.04]
Week 24	85	55.2 (19.8)	72	52.3 (20.8)	74	56.0 (21.2)	$X^2(2) = 2.24$ p = 0.326	-2.30 [-7.00, 2.41] p = 0.339	-0.14 [-0.44, 0.15]	-3.80 [-8.87, 1.26] p = 0.141	-0.24 [-0.56, 0.08]
R-GPTS											

Baseline	92	43.8 (17.5)	90	42.6 (17.9)	92	40.3 (18.4)					
Week 6	66	32.9 (20.9)	62	31.6 (21.3)	57	35.2 (22.7)	$X^2(2) = 4.6$ p = 0.086	-3.66 [-8.69, 1.37] p = 0.154	-0.20 [-0.49, 0.08]	-5.11 [-9.96, -0.26] *p = 0.039*	-0.29 [-0.56, -0.01]
Week 12	63	28.5 (19.0)	49	33.0 (23.0)	47	34.0 (23.1)	$X^2(2) = 5.12$ p = 0.077	-5.95 [-11.43, -0.47] *p = 0.033*	-0.33 [-0.64, -0.03]	-5.06 [-10.83, 0.70] p = 0.085	-0.28 [-0.60, 0.04]
Week 18	65	29.1 (20.6)	59	30.7 (23.2)	56	32.1 (22.0)	$X^2(2) = 1.75$ p = 0.416	-2.46 [-8.05, 3.13] p = 0.389	-0.14 [-0.45, 0.17]	-3.67 [-9.23, 1.89] p = 0.196	-0.20 [-0.52, 0.11]
Week 24	64	27.8 (20.9)	55	32.0 (22.8)	55	32.5 (22.4)	$X^2(2) = 4.59$ p = 0.101	-5.51 [-10.86, -0.17] *p = 0.043*	-0.31 [-0.61, -0.01]	-3.85 [-9.06, 1.36] p = 0.148	-0.21 [-0.51, 0.08]
PANSS											
Baseline	92	4.1 (0.8)	90	4.1 (0.8)	92	4.1 (0.8)					
Week 6	68	3.6 (1.0)	69	3.3 (1.2)	66	3.7 (1.2)	$X^2(2) = 3.74$ p = 0.154	-0.03 [-0.33, 0.27] p = 0.822	-0.04 [-0.42, 0.33]	-0.29 [-0.61, 0.04] p = 0.083	-0.36 [-0.77, 0.05]
Week 12	71	3.3 (1.2)	58	3.2 (1.2)	57	3.5 (1.2)	$X^2(2) = 2.68$ p = 0.261	-0.06 [-0.39, 0.28] p = 0.736	-0.07 [-0.49, 0.35]	-0.28 [-0.63, 0.07] p = 0.116	-0.35 [-0.78, 0.09]
Week 18	68	3.0 (1.2)	57	3.1 (1.3)	59	3.2 (1.3)	$X^2(2) = 0.88$ p = 0.643	-0.16 [-0.51, 0.20] p = 0.386	-0.19 [-0.63, 0.25]	-0.14 [-0.51, 0.22] p = 0.447	-0.18 [-0.63, 0.28]

Week 24	70	3.0 (1.2)	56	3.0 (1.2)	65	3.3 (1.3)	$X^2(2) = 1.49$ p = 0.475	-0.20 [-0.57, 0.16] p = 0.279	-0.25 [-0.71, 0.20]	-0.18 [-0.53, 0.16] p = 0.297	-0.23 [-0.66, 0.20]
PWQ											
Baseline	92	10.1 (4.9)	90	10.5 (4.8)	92	9.8 (4.9)					
Week 6	68	8.8 (5.1)	68	8.9 (5.3)	68	8.9 (5.5)	$X^2(2) = 0.89$ p = 0.641	-0.16 [-1.38, 1.05] p = 0.791	-0.03 [-0.28, 0.22]	-0.58 [-1.84, 0.68] p = 0.367	-0.12 [-0.38, 0.14]
Week 12	71	7.3 (5.5)	58	8.1 (5.5)	55	8.1 (5.6)	$X^2(2) = 0.81$ p = 0.134	-0.34 [-1.59, 0.91] p = 0.594	-0.07 [-0.33, 0.19]	-0.64 [-2.04, 0.77] p = 0.374	-0.13 [-0.42, 0.16]
Week 18	71	6.8 (5.1)	58	6.8 (5.2)	61	7.7 (5.8)	$X^2(2) = 4.01$ p = 0.134	-0.39 [-1.66, 0.89] p = 0.551	-0.08 [-0.34, 0.18]	-1.27 [-2.58, 0.04] p = 0.058	-0.26 [-0.53, 0.01]
Week 24	73	7.3 (5.1)	58	6.0 (4.8)	62	8.1 (5.9)	$X^2(2) = 10.64$ **p = 0.005**	-0.65 [-1.93, 0.64] p = 0.325	-0.13 [-0.40, 0.13]	-2.19 [-3.54, -0.84] **p = 0.001**	-0.45 [-0.73, -0.17]
HADS (Anxiety)											
Baseline	92	12.8 (4.0)	90	12.9 (4.7)	92	12.2 (4.6)					
Week 6	69	12.6 (4.8)	69	11.3 (4.6)	68	11.7 (4.3)	$X^2(2) = 3.23$ p = 0.199	-0.09 [-1.16, 0.98] p = 0.870	-0.02 [-0.26, 0.22]	-0.87 [-1.90, 0.15] p = 0.095	-0.20 [-0.43, 0.03]
Week 12	71	10.8 (4.6)	58	11.1 (4.6)	55	10.3 (5.0)	$X^2(2) = 0.16$ p = 0.921	0.02 [-1.22, 1.25] p = 0.979	0.00 [-0.28, 0.28]	-0.22 [-1.53, 1.09] p = 0.743	-0.05 [-0.35, 0.25]

Week 18	71	10.8 (5.0)	59	10.4 (4.5)	61	10.5 (4.8)	$X^2(2) = 1.71$ p = 0.425	0.38 [-0.83, 1.60] p = 0.537	0.09 [-0.19, 0.36]	-0.43 [-1.69, 0.83] p = 0.504	-0.10 [-0.38, 0.19]
Week 24	72	11.0 (5.3)	56	9.7 (4.7)	61	10.6 (4.9)	$X^2(2) = 6.29$ *p = 0.043*	-0.08 [-1.33, 1.17] p = 0.901	-0.02 [-0.30, 0.27]	-1.48 [-2.77, -0.19] *p = 0.025*	-0.33 [-0.63, -0.04]
HADS (Depression)											
Baseline	92	9.7 (4.6)	90	10.2 (4.3)	92	9.6 (4.1)					
Week 6	69	9.3 (4.9)	69	8.5 (4.4)	68	9.2 (4.6)	$X^2(2) = 3.68$ p = 0.159	-0.12 [-1.15, 0.91] p = 0.817	-0.03 [-0.27, 0.21]	-0.97 [-2.04, 0.11] p = 0.078	-0.22 [-0.47, 0.03]
Week 12	71	8.1 (4.4)	58	8.2 (4.7)	55	8.7 (4.4)	$X^2(2) = 1.57$ p = 0.457	-0.03 [-1.15, 1.08] p = 0.951	-0.01 [-0.27, 0.25]	-0.73 [-2.02, 0.56] p = 0.266	-0.17 [-0.47, 0.13]
Week 18	71	7.9 (5.3)	59	8.4 (4.9)	61	8.9 (5.4)	$X^2(2) = 0.33$ p = 0.849	-0.32 [-1.69, 1.05] p = 0.647	-0.07 [-0.39, 0.24]	-0.40 [-1.85, 1.04] p = 0.583	-0.09 [-0.43, 0.24]
Week 24	72	8.2 (5.7)	56	7.4 (4.5)	61	8.5 (5.6)	$X^2(2) = 6.23$ *p = 0.044*	-0.20 [-1.62, 1.23] p = 0.784	-0.05 [-0.38, 0.28]	-1.52 [-2.88, -0.16] *p = 0.028*	-0.35 [-0.67, -0.04]
SRT (Bias)											
Baseline	92	-0.1 (0.9)	90	0.0 (0.9)	92	-0.1 (0.9)					
Week 6	66	-0.5 (0.9)	62	-0.4 (1.1)	54	-0.1 (1.0)	$X^2(2) =$ 24.93 **p = 0.001**	-0.35 [-0.54, -0.16] **p < 0.001**	-0.39 [-0.60, -0.17]	-0.48 [-0.67, -0.28] **p < 0.001**	-0.53 [-0.75, -0.31]

Week 12	62	-0.4 (1.0)	47	-0.6 (0.9)	46	-0.1 (1.1)	$X^2(2)$ = 47.13 **p = 0.001**	-0.31 [-0.49, -0.14] **p < 0.001**	-0.35 [-0.55, -0.15]	-0.62 [-0.80, -0.44] **p < 0.001**	-0.69 [-0.89, -0.49]
Week 18	64	-0.5 (1.1)	56	-0.5 (1.0)	56	-0.1 (0.9)	$X^2(2)$ = 28.06 **p = 0.001**	-0.29 [-0.48, -0.10] **p = 0.003**	-0.33 [-0.54, -0.11]	-0.47 [-0.65, -0.30] **p < 0.001**	-0.53 [-0.73, -0.33]
Week 24	63	-0.6 (1.0)	55	-0.5 (1.0)	55	-0.1 (1.0)	$X^2(2)$ = 25.76 **p = 0.001**	-0.30 [-0.50, -0.09] **p = 0.005**	-0.33 [-0.56, -0.10]	-0.51 [-0.70, -0.31] **p < 0.001**	-0.57 [-0.78, -0.35]
SST (Bias)											
Baseline	92	43.9 (27.2)	90	47.2 (27.0)	92	43.1 (24.3)					
Week 6	66	35.8 (29.0)	62	31.2 (27.7)	57	43.5 (26.4)	$X^2(2) = 8.92$ *p = 0.012*	-6.32 [-15.23, 2.59] p = 0.164	-0.24 [-0.58, 0.10]	-14.40 [-23.87, -4.92] **p = 0.003**	-0.55 [-0.91, -0.19]
Week 12	62	33.1 (30.5)	48	31.5 (30.8)	47	41.4 (29.0)	$X^2(2) = 4.4$ p = 0.111	-6.30 [-16.31, 3.72] p = 0.218	-0.24 [-0.62, 0.14]	-11.35 [-21.98, -0.73] *p = 0.036*	-0.43 [-0.84, -0.03]
Week 18	62	33.3 (26.7)	55	34.1 (31.2)	55	36.1 (30.6)	$X^2(2) = 0.93$ p = 0.629	-2.79 [-12.38, 6.79] p = 0.568	-0.11 [-0.47, 0.26]	-5.19 [-15.76, 5.38] p = 0.336	-0.20 [-0.60, 0.21]
Week 24	57	30.1 (31.7)	51	31.3 (28.8)	46	34.3 (32.1)	$X^2(2) = 1.09$ p = 0.579	-1.62 [-11.89, 8.64] p = 0.757	-0.06 [-0.45, 0.33]	-5.20 [-15.36, 4.95] p = 0.315	-0.20 [-0.59, 0.19]
WEMWBS											
Baseline	92	37.9 (10.4)	90	37.1 (10.2)	92	37.5 (9.0)					

Week 6	69	37.6 (11.1)	69	40.1 (10.2)	68	38.6 (10.5)	$X^2(2) = 3.67$ p = 0.160	-0.77 [-2.90, 1.36] p = 0.480	-0.08 [-0.29, 0.14]	1.42 [-0.86, 3.71] p = 0.222	0.14 [-0.09, 0.38]
Week 12	71	41.3 (11.9)	58	40.3 (9.3)	55	39.6 (11.4)	$X^2(2) = 0.23$ p = 0.892	0.11 [-2.69, 2.92] p = 0.938	0.01 [-0.27, 0.30]	0.61 [-2.11, 3.32] p = 0.661	0.06 [-0.21, 0.34]
Week 18	70	42.4 (13.1)	59	41.7 (11.1)	60	40.1 (10.7)	$X^2(2) = 0.23$ p = 0.892	0.23 [-2.38, 2.84] p = 0.863	0.02 [-0.24, 0.29]	0.68 [-2.13, 3.48] p = 0.636	0.07 [-0.22, 0.35]
Week 24	71	43.0 (12.8)	55	43.2 (10.4)	61	39.9 (11.7)	$X^2(2) = 4.76$ p = 0.093	2.06 [-0.77, 4.88] p = 0.154	0.21 [-0.08, 0.50]	3.13 [0.28, 5.98] *p = 0.031*	0.32 [0.03, 0.61]
QPR											
Baseline	92	31.8 (12.3)	90	30.6 (11.6)	92	29.4 (10.4)					
Week 6	69	31.9 (13.0)	69	34.9 (10.6)	68	31.3 (11.2)	$X^2(2) = 8.69$ *p = 0.013*	-0.81 [-3.04, 1.41] p = 0.474	-0.07 [-0.26, 0.12]	2.44 [0.27, 4.60] *p = 0.027*	0.21 [0.02, 0.40]
Week 12	71	36.4 (12.8)	58	34.8 (11.6)	55	32.0 (11.2)	$X^2(2) = 0.73$ p = 0.696	1.08 [-1.42, 3.59] p = 0.396	0.09 [-0.12, 0.31]	0.69 [-2.25, 3.63] p = 0.644	0.06 [-0.20, 0.32]
Week 18	70	36.6 (14.0)	59	35.3 (11.6)	61	33.9 (12.1)	$X^2(2) = 0.21$ p = 0.900	-0.55 [-3.27, 2.17] p = 0.692	-0.05 [-0.29, 0.19]	-0.64 [-3.88, 2.60] p = 0.697	-0.06 [-0.34, 0.23]
Week 24	72	36.4 (13.8)	56	37.9 (11.6)	62	33.0 (12.6)	$X^2(2) = 3.8$ p = 0.150	0.64 [-2.24, 3.52] p = 0.662	0.06 [-0.20, 0.31]	2.59 [-0.21, 5.40] p = 0.070	0.23 [-0.02, 0.47]

ReQoL											
Baseline	92	17.8 (8.5)	90	18.2 (7.4)	92	17.6 (7.0)					
Week 6	69	18.6 (8.5)	69	20.5 (8.1)	68	19.3 (7.6)	$X^2(2) = 0.87$ p = 0.648	-0.47 [-2.15, 1.20] p = 0.580	-0.06 [-0.28, 0.16]	0.37 [-1.44, 2.18] p = 0.685	0.05 [-0.19, 0.29]
Week 12	71	20.8 (8.8)	58	21.1 (7.8)	55	20.7 (8.5)	$X^2(2) = 1.15$ p = 0.562	-1.11 [-3.18, 0.96] p = 0.294	-0.15 [-0.42, 0.13]	-0.37 [-2.35, 1.61] p = 0.715	-0.05 [-0.31, 0.21]
Week 18	70	22.4 (9.6)	59	21.9 (7.9)	61	20.9 (7.7)	$X^2(2) = 0.17$ p = 0.918	0.32 [-1.48, 2.13] p = 0.725	0.04 [-0.19, 0.28]	0.02 [-1.95, 1.98] p = 0.986	0.00 [-0.26, 0.26]
Week 24	71	21.6 (9.3)	55	23.1 (8.5)	62	20.3 (8.2)	$X^2(2) = 3$ p = 0.223	0.63 [-1.34, 2.61] p = 0.530	0.08 [-0.18, 0.34]	1.81 [-0.25, 3.88] p = 0.085	0.24 [-0.03, 0.51]

Outcomes include clinical symptoms (e.g., paranoia, anxiety, depression), cognitive bias (SRT, SST), and recovery measures (process of recovery, quality of life, well-being). PS: Paranoia Scale; R-GPTS: Green Paranoid Thoughts Scale Revised; PWQ: Paranoia Worries Questionnaire; HADS: Hospital Anxiety and Depression Scale; WEMWBS: Warwick-Edinburgh Mental Wellbeing Scale; QPR: Questionnaire about the Process of Recovery; ReQoL: Recovering Quality of Life. Our statistical analysis plan specified a primary outcome alpha level of 0.05; a secondary outcome alpha level of 0.01 for 5 key outcomes (SRT, SST, GPTS, HADS, PWQ) and for the remaining variables, explorative statistics and estimated treatment effects without formal hypothesis testing. Bold denotes p values meeting these respective thresholds, italic indicates a trend.

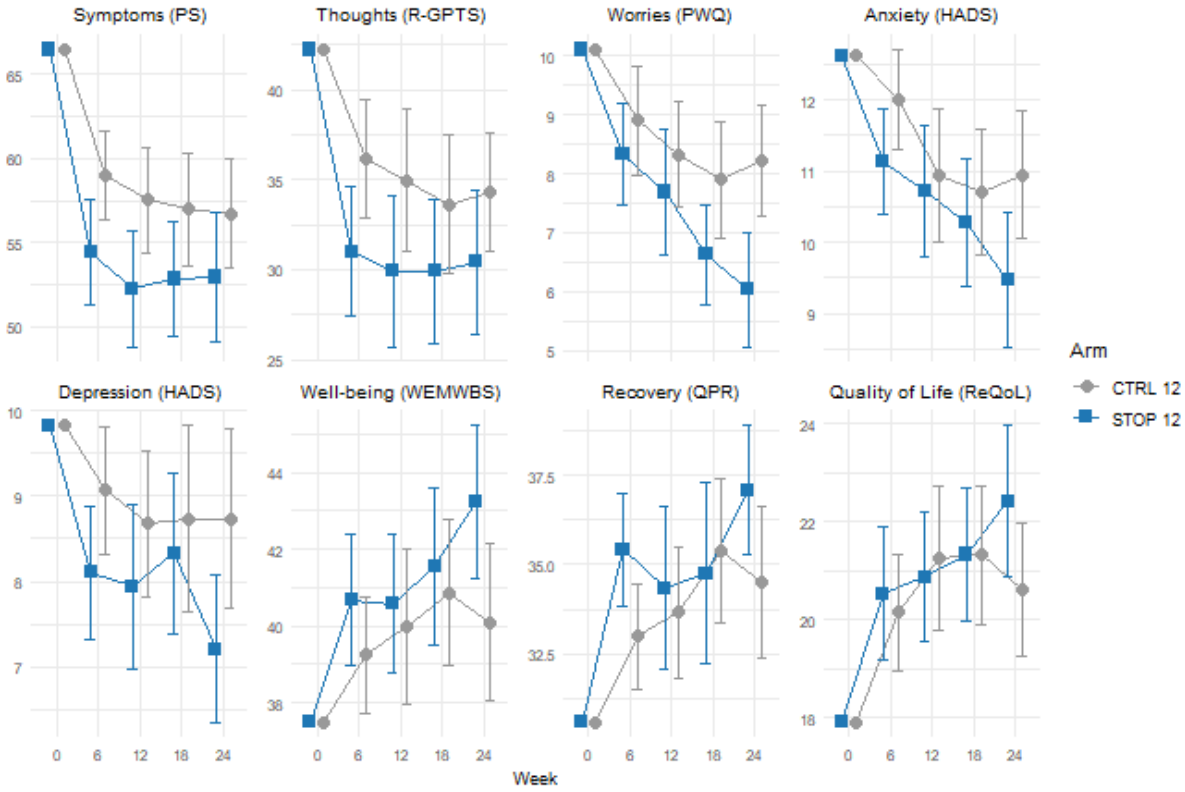


Fig. 2 Predicted means over time for primary and secondary outcomes comparing STOP 12 and CTRL 12. Error bars represent 95% confidence intervals. PS: Paranoia Scale; R-GPTS: Green Paranoid Thoughts Scale Revised; PWQ: Paranoia Worries Questionnaire; HADS: Hospital Anxiety and Depression Scale; WEMWBS: Warwick-Edinburgh Mental Wellbeing Scale; QPR: Questionnaire about the Process of Recovery; ReQoL: Recovering Quality of Life.

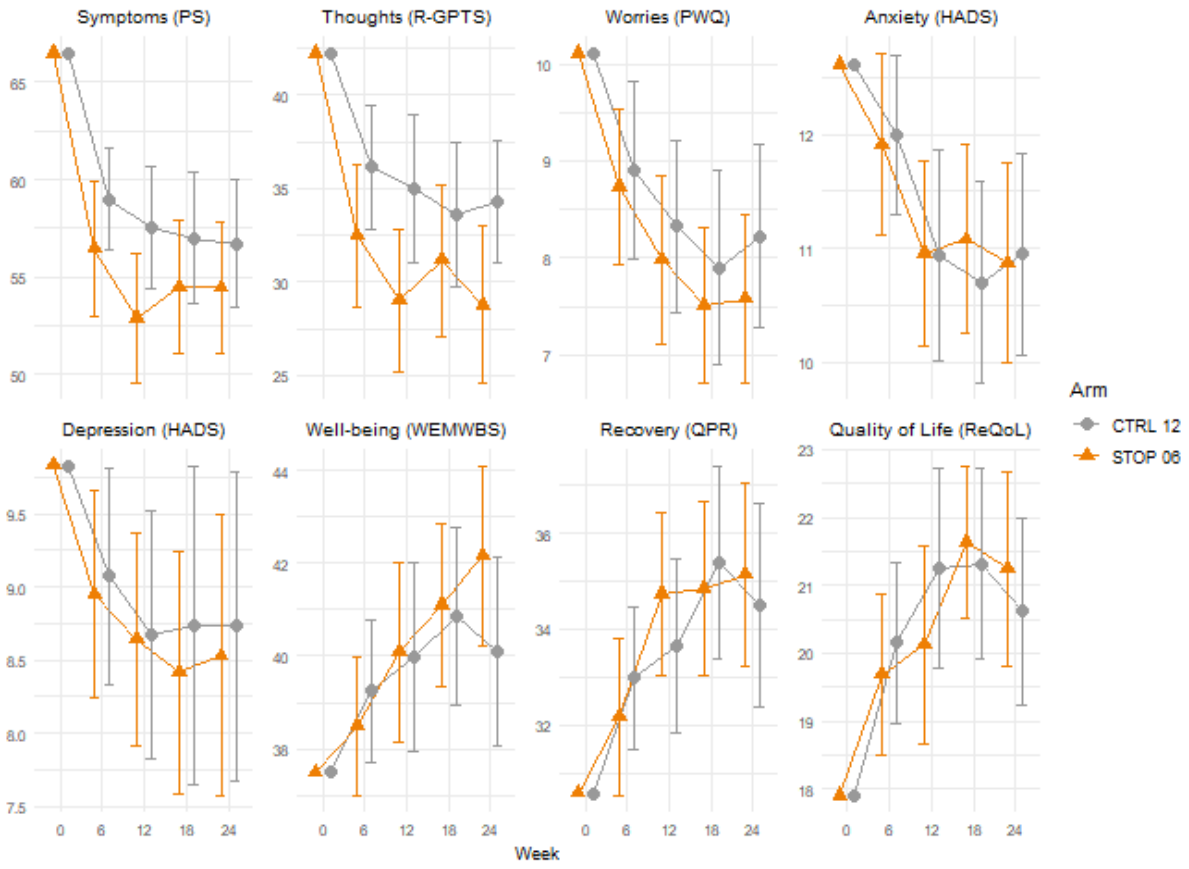


Fig. 3. Predicted means over time for primary and secondary outcomes comparing STOP 6 and CTRL 12. Error bars represent 95% confidence intervals. PS: Paranoia Scale; R-GPTS: Green Paranoid Thoughts Scale Revised; PWQ: Paranoia Worries Questionnaire; HADS: Hospital Anxiety and Depression Scale; WEMWBS: Warwick-Edinburgh Mental Wellbeing Scale; QPR: Questionnaire about the Process of Recovery; ReQoL: Recovering Quality of Life.

Secondary Outcomes

Table 2 and Figures 2 and 3 also show the treatment effect estimates across clinical and recovery measure secondary outcomes.

Mechanism Outcomes (Bias)

The **SRT** task bias reduction was evident by Week 6 continuing through to Week 24 (all joint test p's < 0.001). In both STOP 12 (ES = -0.48, 95% CI: -0.67, -0.28) and STOP 6 (ES = -0.39 95% CI: -0.60, -0.17), increasing to ES = -0.69 (95% CI: -0.89, -0.49) for STOP 12 but not

increasing for STOP 6 (ES = -0.35) with this difference still present at week 24 (STOP 12 ES = -0.57 (95% CI: -0.73, -0.33) vs STOP 6 ES = -0.33 (95% CI: -0.54, -0.11)).

The **SST** showed an initial bias reduction at 6 weeks (χ^2 (2) = 8.92, p = 0.012), STOP 12 ES = -0.55 (95% CI: -0.91, -0.19) and STOP 6 ES = -0.24 (95% CI: -0.58, -0.10). By 24 weeks differences were smaller (STOP 12 ES = -0.2 and STOP 6 ES = -0.06).

Clinical Outcomes

Paranoia

R-GPTS broadly showed the same pattern of results to the paranoia scale, both for the total score and the reference and paranoia subscales, total score ES -0.21 to -0.35 at weeks 12 and 24 for both STOP 12 and STOP 6 vs CTRL 12, corresponding to differences ranging from –3.85 to –5.95 points.

PANSS scores were lower for STOP 6 and STOP 12 participants than control at 24 weeks (STOP 6 ES = -0.25 95% CI: -0.71, 0.20 and STOP 12 -0.23 95% CI: -0.66, 0.20), equivalent to point drops of 0.2 and 0.18 on the PANSS scale (single item). The **PWQ** showed a substantial delayed response with ES for STOP 6 and STOP 12 relatively small at week 12 (ES's -0.07 and -0.13) but by week 24 STOP 12 showed a significant reduction relative to CTRL 12 (ES = -0.45 95% CI: -0.73, -0.17, p = 0.001). This corresponded to an adjusted group difference of 2.19 points. ES for STOP 6 remained modest, although consistent, with a 0.65-point decrease in symptoms (ES = -0.31).

Anxiety and Depression

Anxiety showed a greater reduction in the STOP 12 condition relative to CTRL 12 and STOP 6 across the trial. At week 24, the ES of the difference was -0.33 (95% CI: -0.63, -0.04) relative to CTRL 12 corresponding to a 1.48-point reduction on the anxiety sub-scale of HADS. Similarly, there was a reduction in **depression** scores, ES = -0.35 (95% CI -0.65, -0.04) corresponding to a 1.52-point drop (STOP 12 vs CTRL 12).

Recovery Outcomes

The 3 recovery outcomes (WEMWBS, QPR, and ReQoL) showed the same pattern where outcomes improved post end of treatment and in particular for the STOP 12 condition. Recovery was consistently better for STOP 12 at 24 weeks relative to CTRL 12, WEMWBS: 3.13 points difference ES = 0.32 (95% CI: 0.03, 0.61), QPR: 2.59 point difference, ES = 0.23, 95% CI: (-0.02, 0.47) and ReQoL 1.81 point difference ES = 0.24, 95% CI: (-0.03, 0.51). ES for STOP 6 ranged from 0.21, 0.06 to 0.08 for WEMWBS, QPR and ReQoL respectively.

Figure 4 provides a visual summary of standardized effect sizes across all outcomes at weeks 12 and 24.

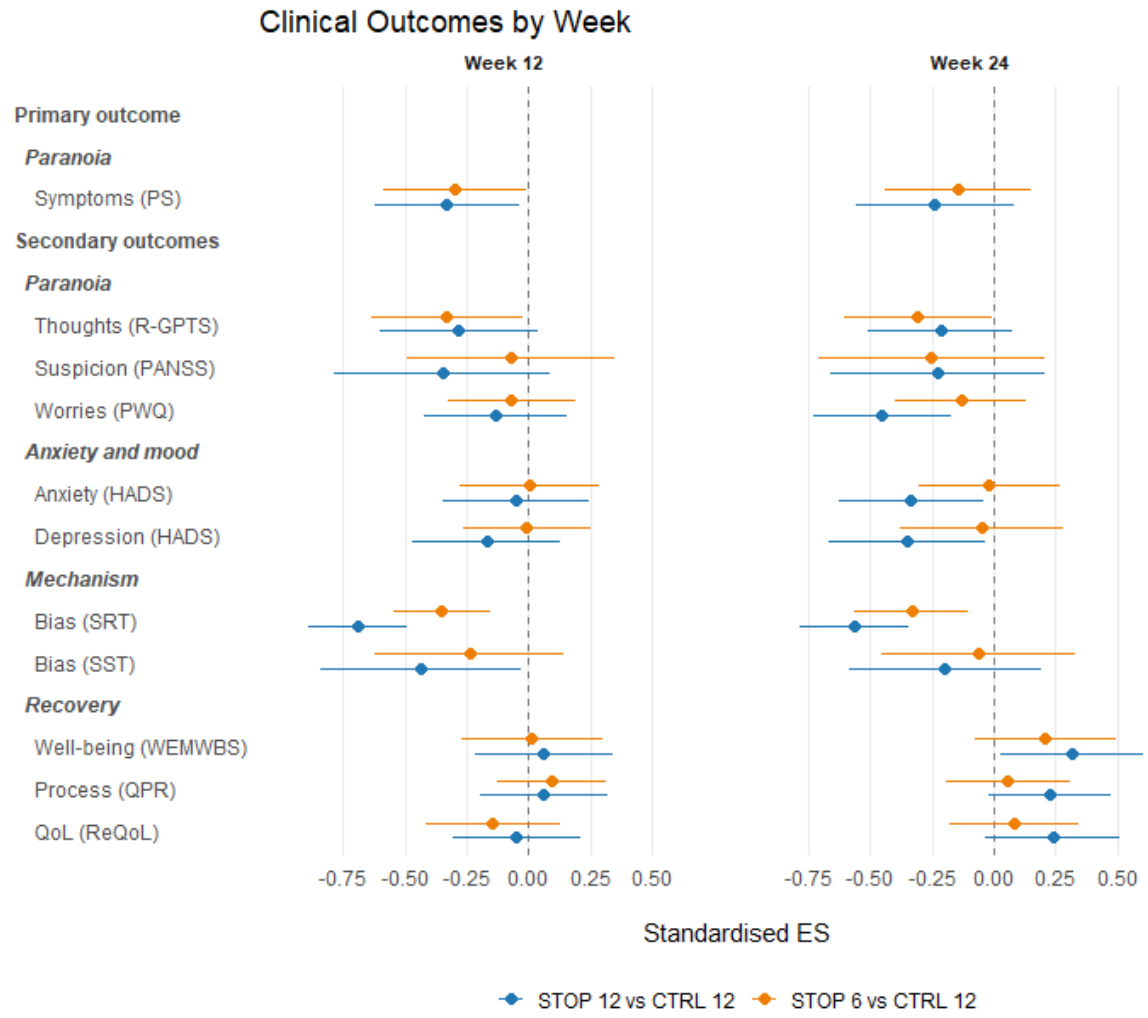


Figure 4. Standardized effect sizes (ES) and 95% confidence intervals across clinical, bias, and recovery outcomes at Week 12 and Week 24. Effect sizes are shown for comparisons between STOP 12-week and control (blue) and STOP 6-week and control (orange) groups. More positive values indicate a beneficial effect of the STOP intervention relative to control, with the exception of recovery measures which are reversed. Outcomes include clinical symptoms (e.g., paranoia, anxiety, depression), cognitive bias (SRT, SST), and recovery measures (process of recovery, quality of life, well-being). PS: Paranoia Scale; R-GPTS: Green Paranoid Thoughts Scale Revised; PWQ: Paranoia Worries Questionnaire; HADS: Hospital Anxiety and Depression Scale; WEMWBS: Warwick-Edinburgh Mental Wellbeing Scale; QPR: Questionnaire about the Process of Recovery; ReQoL: Recovering Quality of Life.

Treatment Adherence

Target treatment adherence was set at 75% and was predefined as meeting two criteria: (1) session completion, defined as providing responses for at least 75% of trials within a session,

and (2) completion of at least 50% of scheduled sessions[37]. Based on these criteria, 85% (232/274) of the sample were adherent, 10 % higher than the target.

Both the intervention and comparator were delivered as intended. During the trial 38 protocol violations were recorded. Twenty-two were data collection-related, such as missing data points due to staff absences, incorrect data entry, or participants completing a measure twice. Nine were due to unmasking, through for example participants sharing information about their sessions. Seven were enrolment-related, such as sessions becoming available on a participant's phone before randomisation or participants missing session one due to procedural error.

Concomitant care

Throughout the post-randomisation period, participants in all three arms continued with their Treatment as Usual (TAU; see Table S1). TAU consisted of psychological therapies not excluded by the eligibility criteria and psychotropic medications. Twenty-four participants (24/274; 8.76%) were receiving psychological therapy during their participation in the trial, while 223 participants (223/274; 81.39%) were receiving psychotropic medication. The most commonly prescribed psychotropic medication was antipsychotics (169/274; 61.68%), followed by antidepressants (119/274; 43.43%).

Safety

A total of 243 out of 274 participants (88.69%) experienced at least one adverse event (AE) during the trial. Across all participants, 1,511 AEs were reported, distributed as follows: 492 (33.22%) in the 6-week arm, 521 (35.18%) in the 12-week arm, and 498 (33.63%) in the control arm. In the 6-week arm, 81 of 92 participants (88.04%) experienced an AE; in the 12-week arm, 78 of 90 participants (86.67%); and in the control arm, 83 of 92 participants

(90.22%). Thirteen serious adverse events (SAEs) (to be reported in more detail in the STOP safety paper; published separately) were reported during the trial; all were deemed unrelated to the intervention. Of these, 7 (53.85%) occurred in the 6-week arm, 4 (30.77%) in the 12-week arm, and 2 (15.38%) in the control arm.

User satisfaction

Table S2 presents participant feedback on the acceptability and usability of the STOP intervention using the Ben-Zeev instrument[58]. The majority of participants reported a positive experience, with over 94.65% stating that the content was easy to read. They were satisfied with the system's ease of use (91.44%) and found it straightforward to learn (89.84%). A similar proportion felt comfortable using STOP (89.84%) and believed that most people would be able to learn how to use it quickly (89.3%). Confidence in using the intervention was high (82.89%), and most participants felt that its various components worked well together (76.47%). In addition, more than 83% found it easy to locate relevant information, and nearly 77% said they would recommend STOP to others. Just over half (57.22%) felt that the intervention helped them manage their symptoms. Only a minority encountered difficulties: 12.83% described it as complicated, 6.42% felt they needed to learn a lot before getting started, 7.49% found it awkward to use, and 11.23% reported needing technical support. Overall, the findings indicate that most participants found STOP to be both accessible and user-friendly.

Discussion

This study investigated the effectiveness, in reducing symptoms of paranoia, of two different doses of the STOP app (6-weeks; STOP 6 and 12-week; STOP 12) compared to a placebo control app. Adherence was high (85%), with no intervention-related serious adverse

events. All three arms showed a reduction in paranoia on the primary outcome (PS) compared to baseline levels. Both STOP doses were effective in reducing paranoid bias compared to control (ds = -.29 to -.62) after 12 weeks. The 12-week dose showed the strongest symptom effects, with significant reductions in paranoia on the primary outcome compared to control post treatment (d = -0.33, p = 0.026), although the primary endpoint (24 weeks) was not significant (d = -0.24, p = 0.141). The pattern of results on the primary outcome suggests the presence of a treatment signal that warrants further work to strengthen and prolong the efficacy of STOP.

Statistical significance was variable across other outcomes and timepoints, although the direction of effects remained consistent, further supporting the above conclusion, particularly for the 12-session version of STOP. Significant improvements were observed at 24 weeks in paranoia-related worry (d = -0.45, p = 0.001), with smaller (but significant) gains in anxiety, depression and some recovery measures. Paranoia outcomes measured by the R-GPTS aligned with the primary findings, with small-to-moderate effect sizes (0.2 to 0.35) observed at both 12 and 24 weeks for STOP 6 and STOP 12 compared to the control. Similarly, PANSS scores were lower in both STOP groups at 24 weeks, with odds ratios from an ordinal regression model indicating a reduced likelihood of higher symptom severity—although wide confidence intervals indicated uncertainty in the differences. These findings suggest a possible effect on broader symptom domains that warrants further exploration.

Paranoia-related worries, as measured by the PWQ, showed a delayed but strong treatment response. Initial effect sizes were small at 12 weeks but STOP 12 participants experienced a significantly greater reduction in paranoia worries by 24 weeks (ES = -0.45, p = 0.001), a result that remained significant even after correcting for multiple comparisons. This suggests that STOP 12 may be particularly effective at reducing persistent cognitive-affective

features of paranoia over time. Anxiety and depression symptoms followed a similar trajectory: by 24 weeks, STOP 12 participants had significantly lower scores on both subscales compared to the control group (anxiety ES = -0.33; depression ES = -0.35), each corresponding to a 1.5-point reduction. These effects support the notion that the intervention may also alleviate broader emotional distress associated with paranoia.

Recovery outcomes were also strongest for the STOP 12 intervention. All three indicators (WEMWBS, QPR, and ReQoL) improved over time, with the most consistent gains observed at the 24-week follow-up in the STOP 12 group. These findings suggest that the intervention may contribute to symptom reduction and improvements in wellbeing and perceived recovery.

Mechanistic outcomes related to bias reduction provided support for the theoretical basis of the intervention. On the SRT task, both STOP arms showed significant reductions in interpretation bias by Week 6 (joint test $p < 0.001$), with STOP 12 participants exhibiting the greatest and most sustained improvement through to Week 24 (ES = -0.57). This suggests that the 12-session intervention may be most effective in shifting the core cognitive processes hypothesised to maintain paranoid thinking. In contrast, bias reduction effects on the SST task were more limited. While both STOP groups showed significant improvement at Week 6, these effects diminished by 24 weeks, largely due to improvements observed in the control group during the follow-up period. One plausible explanation lies in what the two tasks measure: the SRT is thought to index more explicit, deliberative reasoning (‘thinking slow’), whereas the SST captures more implicit, rapid judgments (‘thinking fast’). Because STOP trains the deliberate generation of alternative, non-threatening interpretations, greater and more durable change on the SRT is consistent with its proposed mechanism.

Taken together, these findings suggest that digital interventions targeting interpretation bias may offer a promising and accessible way of reducing paranoia and improving related outcomes in real-world clinical populations. Importantly, effects were demonstrated relative to a placebo control condition, which itself showed improvements over time, possibly due to regular contact with a researcher and structured engagement. This provides a more rigorous test than a waitlist control, which would not be expected to show similar levels of change. The results suggest that a longer, 12-session format is more appropriate than a shorter 6 week format, especially in producing improvements in paranoia, affective symptoms, recovery, and cognitive bias. While not all effects reached statistical significance, the consistency of the pattern across measures and timepoints, coupled with significant results in key domains even after correction for multiple comparisons, supports the clinical relevance of this approach. A valuable next step for future research and development would be to explore the benefits of personalizing the intervention, for example, adapting content to individual needs or contexts such as employment concerns, parenting, or other dominant life themes which may help strengthen engagement and enhance clinical effectiveness. The opportunities offered by AI could make personalization especially effective, for example by using a real-time approach in which last week's challenges are transformed into next week's therapeutic intervention.

Importantly, the intervention was delivered entirely without therapist input, and the high rates of adherence and acceptability provide support for its scalability. This delivery model offers substantial potential to reach underserved populations and reduce demands on clinical services. The STOP app may particularly appeal to people with paranoia who may not want, or not feel ready for a talking therapy, such as CBT for psychosis as the app does not require anyone to talk about or disclose their own personal worries or concerns.

Several limitations should be acknowledged. The sample was restricted to English-speaking individuals with moderate to severe paranoia and interpretation bias, potentially limiting generalizability to those with milder, fluctuating, or more severe presentations. Individuals with extreme levels of paranoia (PANSS item 6 score of 7) or high suicide risk were excluded, meaning the intervention's relevance to the most severe or complex cases remains unknown. The intervention was designed as a fully self-administered digital tool without therapist input, which may not suit those needing interpersonal or motivational support. While this format enhances scalability, successful implementation will likely require digital navigators or low-intensity support, particularly for individuals with limited digital literacy or motivational barriers, as highlighted in implementation research on other digital mental health interventions. Offering a mobile phone to participants is also important in research testing an app-based intervention for paranoia, as asking participants to use their personal phone may be perceived as intrusive or raise concerns about data security. We therefore offered participants the option of using a study phone (including a pre-paid data package) to mitigate digital exclusion for people who could not afford their own smartphone; over 1/3 took this up. The majority of participants (77.8%) were referred from secondary care services, suggesting that many may have had more severe or complex presentations of paranoia, which is important to consider when interpreting outcomes. Finally, reliance on self-reported measures for both primary and secondary outcomes introduces the possibility of response biases, such as social desirability or expectancy effects.

In conclusion, this study provides evidence that a self-guided digital intervention targeting interpretation bias may reduce paranoia and improve related clinical outcomes. These findings support the use of targeted approaches within digital mental health and highlight the potential for scalable technological interventions to expand access to mental health solutions.

Methods

Study design

This is a double-blind, superiority, three-arm RCT. Participants were recruited from across England and Wales. Participants were randomly allocated to a 6-session intervention group; a 12-session intervention group; or a 12-session text reading control. TAU continued for all participants (allocation ratio 1:1:1). The trial was designed to run entirely remotely. The full trial protocol was published elsewhere[38]. This paper follows the CONSORT (Consolidated Standards of Reporting Trials) 2025 checklist.[39,40]

The study was prospectively registered on ISRCTN before recruitment began (ISRCTN17754650; 3/08/2021).

Study intervention

STOP intervention. Participants randomized to the intervention received 40 training items per weekly session on the STOP mobile app (version 1.6). Participants read text inviting paranoid interpretations, then completed missing words and answer questions in a way that encourages alternative, adaptive beliefs about themselves and others (see supplementary material for screenshots of the STOP intervention). Full details of the development of the STOP intervention have been published by Hsu and colleagues [34]. Figure 5 shows example screenshots of STOP.

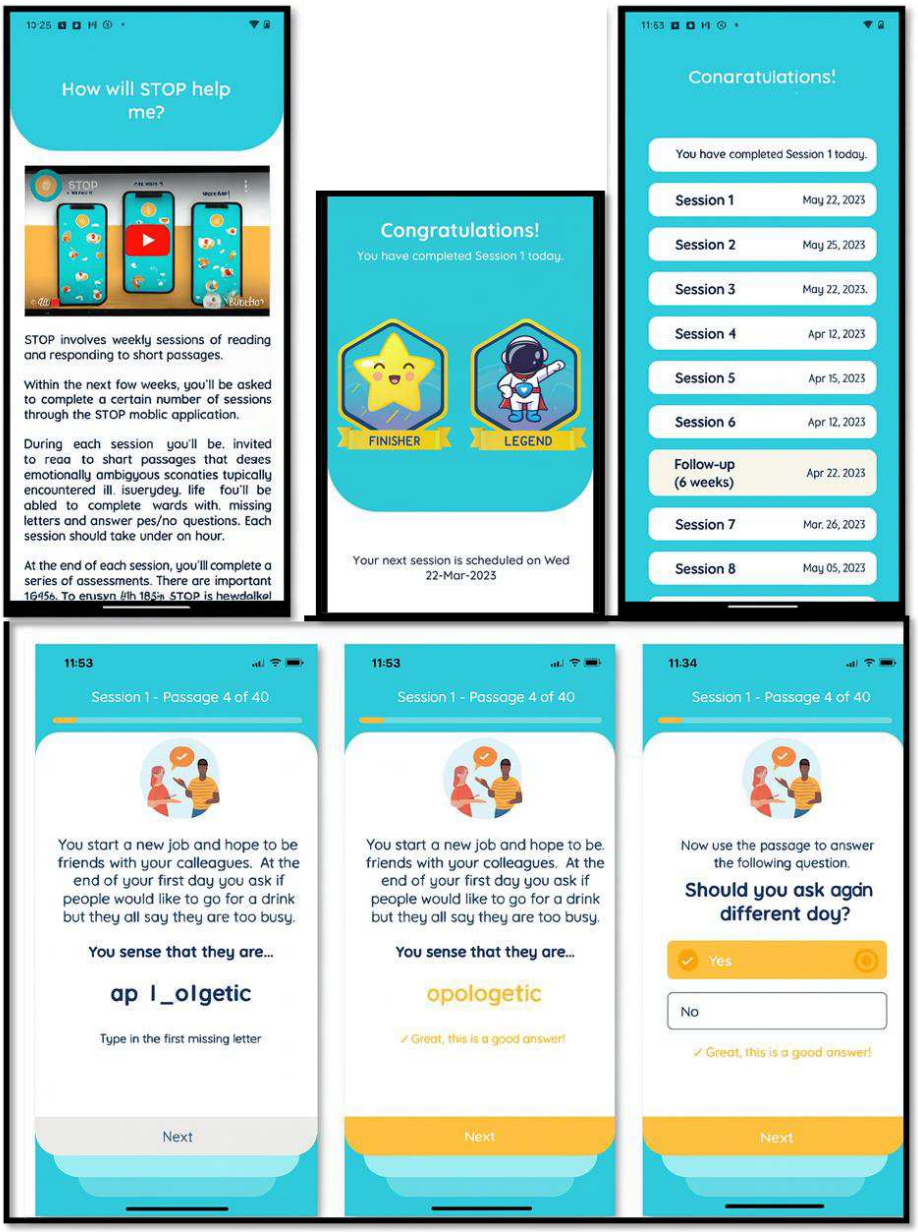


Fig. 5: Example screen shots from the STOP smartphone app. Top panel shows introduction page, badge reward screen and therapeutic session home page. Bottom panel shows an example of one individual scenario from the paranoia bias modification therapy.

In the six-session arm of STOP, users receive only the first 6 sessions of the full 12-session therapy. The first six sessions of the intervention were intentionally designed to include milder, less threatening content reflecting lower levels of paranoia. They also used only the earlier, less extreme verbs to ensure the material was appropriate and manageable for participants during the initial phase of the program.

Text-reading control. Participants randomized to the placebo control condition received 40 control items per session (approximately 40 min per session) on the STOP mobile app. The experience is identical to the intervention condition except the item content that participants see omits the 'active ingredient' (resolution of a potentially triggering, but ambiguous, situation in a benign/non-paranoid manner. Instead, control participants read and respond to factual material or mundane everyday experiences (see supplementary material for screenshots of the control).

Participants

Recruitment

The trial was conducted fully online and had a wide range of recruitment pathways including: 20 NHS Trusts across England; our main UK charity collaborators, the McPin Foundation; MQ Mental Health Research (a UK charity); other third sector organisations accessed through McPin; institutional research registers (lists of individual patients who have consented in advance to be contacted about research participation) , and self-referral through the study website. Out of 1017 referrals 569 did not proceed to eligibility assessment for the following reasons: 225 lost interest, 239 became uncontactable, 62 became ineligible, and 43 were lost for other reasons. Of the 274 randomized to the study, 70% (191) were from secondary

care, 12% (33) via the McPin charity and 18% (50) from other sources (self-referral, MQ charity recruitment website, personal contacts or King's university counselling service).

During recruitment, participants were introduced to the study and provided with a Participant Information Sheet (PIS) and a consent form (hard copy or online, depending on participant preference). They were given at least 24 hours to consider their decision and ask any questions before deciding whether to join the study. Informed consent was taken before screening due to the significant and necessary data collection that screening involves. Participants' eligibility was established by systematically assessing the inclusion and exclusion criteria described below. Just over one-third of the sample (89/274; 32.48%) opted to use a study phone.

Eligibility criteria

There were no eligibility criteria relating to current or past use of mental health services or diagnosis. There were no digital literacy requirements for participation. Interested individuals who did not own a smartphone were offered one (including a pre-paid data package), and those with no prior experience using a smartphone were provided with face-to-face support to learn how to operate the device, download the app, and navigate its features.

Inclusion criteria:

1. Any clinically significant persecutory or paranoid symptoms, present for at least the preceding month. This was operationalized as a score of 3 ('mild') or more on item 6 of the Positive and Negative Syndrome Scale (PANSS)[41].

2. Displaying an interpretation bias on the 8-item screening version of the similarity ratings task (SRT)[42,43]. Screening bias scores range from +3 to−3. Zero indicates no bias; plus scores indicate a bias favoring paranoid interpretations and minus scores indicate a bias favoring nonparanoid interpretations. A bias score of −2 or greater (i.e.−2 up to+3) was necessary to be selected for the study.

3. If on psychotropic medication, stable on that medication for at least the last 3 months and expected to be so for the study duration.

4. Age 18 years or older.

5. Capacity to consent: if in doubt, assessed using the Capacity Assessment Tool[44].

Exclusion criteria:

1. Severe cognitive impairment.

2. Illiteracy or inability to understand written and spoken English.

3. Major current physical illness (e.g. cancer, heart disease, stroke, dementia).

4. Major substance or alcohol misuse, assessed by CAGE-AID Substance Abuse Screening Tool[45].

5. Currently receiving, or soon due to receive, a psychological intervention or having done so in the last 3 months. A psychological intervention, for the purpose of the study, was defined as five or more consecutive sessions with a trained therapist, either in an individual or group setting, for the purpose of alleviating symptoms arising from paranoia or related issues.

6. Currently taking part in any other interventional research study.

7. Extreme paranoia, i.e. scoring 7 (defined as 'Extreme— a network of systematized persecutory delusions denotes the patient's thinking, social relations and behavior') on item 6 of the PANSS.

8. At high risk of suicide as indicated by the Columbia Suicide Severity Rating Scale (C-SSRS)—Screen Version[46,47].

Patient and Public Involvement

A Lived Experience Advisory Panel (LEAP) contributed to the design and running of the trial. LEAP met quarterly to receive progress updates and provide input, including on: the consent form and participant/carer information sheets; the website; recruitment strategies; access to user networks; publicity and dissemination; and any other issues that arose.

Randomization and blinding

Randomization was at the individual level in the ratio 1:1:1 and was completed in the STOP app automatically. The software development company, Avegen, received advice and input from King's Clinical Trials Unit (KCTU) to follow appropriate procedures. There were no adjustments permitted to randomization once completed. Researchers and participants remained masked to the allocated condition. Block randomization with randomly varying block sizes was used. Randomization stratifiers (measured at screening) were as follows:

1. Study hub (Bath, London)

2. Sex assigned at birth (male, female)

For further details on the trial's randomization process, refer to the trial's protocol.

Procedure

The participant pathway through the trial is shown in Fig. 1. As this was a double blinded trial, participants in all arms underwent the same procedure. All data collection procedures were conducted remotely via a phone call or a video call based on participants' preference. A detailed assessment schedule is provided in the protocol paper[37].

Eligible individuals who chose to take part then met remotely with a trained researcher who administered the baseline assessment; where the following socio-demographic data were collected: age, sex at birth, ethnicity, educational level, employment, living arrangement, relationship status, self-reported dyslexia and age of onset of distressing paranoia. The persuadability subscale of the Multidimensional Iowa Suggestibility Scale (MISS) [48] was also administered at baseline as part of the safety protocol. Participants who scored > 58 (2.5 standard deviations above the general population mean) on this subscale were deemed more vulnerable to the possible negative effects of the intervention by the Medicines and Healthcare products Regulatory Authority (MHRA) and were therefore offered weekly check-in phone calls. The researcher guided the participant on how to find and download the app on the App Store or Google Play and provided the participant with a code allowing them app access.

Participants in all arms received in-app weekly post-session assessments for 12 weeks. These assessments comprise a small subset of the main outcome measures and were completed unsupervised within the STOP app. All participants (intervention and control) receive a fortnightly phone call from researchers, following a pre-prepared script to limit bias. Key objectives of the calls were (i) to check participants well-being, and (ii) record any adverse events using the AE checklist (the development of this checklist is published elsewhere[36]).

Follow-up assessments included a full battery of outcome measures at 6, 12, 18 and 24 weeks. Participants completed four measures in the app (PS, G-PTS, SST, and SRT), while the

remaining measures were completed remotely with the researcher using on an online researcher dashboard. Three items ('By the end of the intervention, how much improvement in your paranoid symptoms do you think will occur?'; 'At this point, how much do you really feel that therapy will help you to reduce your symptoms?' and 'By the end of the study, how much improvement in your symptoms do you really feel will occur?') from the Credibility and Expectancy questionnaire[49] were used to measure participants' expectations of benefiting from taking part in the trial, to assess both equipoise and the role of expectation in predicting improvement.

Ethical & Regulatory considerations

The study was approved by the London-Stanmore Research Ethics Committee (reference: 21/LO/0896). STOP was classified as a Software as a Medical Device (SaMD) at the time of the trial. As a result, it needed, and received, a letter of no objection from the MHRA to conduct the trial. The details of the complexity of achieving this regulatory approval are published elsewhere[36].

Outcomes

Primary Outcome

Mean paranoid symptoms measured using the paranoia scale (PS)[35] at 24 weeks post-randomization. The PS is a 20-item self-report measure designed to assess paranoid ideation in response to everyday events and situations. Participants rate their agreement with each item using a 5-point scale (1=not at all applicable to me to 5=extremely applicable to me). Individual scores are summed, ranging from 20 to 100 with higher scores indicating greater endorsement of

paranoid ideation. The PS has been used widely within clinical research and has good psychometric properties with internal consistency, r=0.84, and test–retest reliability, r=0.70[35].

Secondary outcome measures

Secondary outcomes were measured at baseline, 6, 12, 18, and 24 weeks post-randomization, with the 24-week timepoint serving as the primary endpoint. These included interpretation bias measures at each timepoint and after every session using the SRT[42,43] and the SST [50,51]. Clinical symptoms were measured with the R-GPTS [52], the PANSS [41], item 6 only, the HADS [53], PWQ [54]. Recovery was evaluated using the WEMWBS [55], QPR [56], and ReQoL [57]. In addition, user satisfaction was assessed using the instrument developed by Ben-Zeev et al. [58].

Safety Outcomes

Safety in this trial was assessed using multiple measures. In this paper, we present descriptive statistics of the collected AEs and SAEs across different arms. Further data and analysis of safety will be reported separately.

Statistical analysis and sample size calculation

The study was powered to detect clinically meaningful effects based on variance estimates from a previous feasibility study [32]. With alpha=0.05 and 80% power, recruiting a sample of 77 per group would permit the detection of an 8 point drop on the paranoia scale compared to TAU assuming a standard deviation of 17.5. This magnitude drop would be a 15% reduction in the average score of paranoid patients[59,60] and represents an effect size of d=0.46 (small to medium). We assumed a 15% drop-out leading to a total required sample size of 273 (N=91 per arm), using an independent t-test.

An intention to treat analysis based on a treatment policy estimand was conducted using a linear mixed-model for repeated measures (MMRM) to estimate differences between STOP 6 and CTRL 12 and STOP 12 and CTRL 12 at each time point (6, 12, 18, and 24 weeks), with PS as the outcome and baseline PS score, randomization factors (study site, sex at birth), treatment group, time, and interactions between treatment group and time and baseline PS score and time as fixed effects. To account for the repeated measures an unstructured covariance matrix modelled the dependency between residuals. The primary treatment effect was assessed via a global test for differences between arms, followed by pairwise contrasts at 24 weeks. The alpha error level was set at 0.05, controlling the type 1 error for the two planned contrasts by using the LSD test procedure. This procedure does not require a correction of the familywise error of pairwise comparisons of three groups if the main effect is significant. Given the expected increasing dropout rate over time, all time points were analysed simultaneously using restricted maximum likelihood (REML) estimation to reduce bias and maximize statistical power. Continuous baseline covariates were centred at their means. Standardized effect sizes (ES) were calculated by dividing the treatment group differences by the baseline SD of the outcome (Cohen's d). Secondary outcomes were analysed using the same MMRM framework where normality was assumed.

Following ITT principles, missing data were treated as observed. The primary analysis assumed data were missing at random (MAR) and used REML estimation to account for missingness. Sensitivity analyses explored the impact of missing data. Firstly, baseline predictors of missingness were identified via logistic regression, and significant predictors were incorporated into a multiple imputation model at the imputation stage. Missing scale items were

handled using prorating if fewer than 20% of items were missing; otherwise, they were treated as missing data points.

Data Monitoring

Data were provided for the data monitoring committee (DMC) and trial steering committee (TSC) who met biannually to review study procedures, data quality and safety.

Open Science

Trial Registration*:* ISRCTN registry, ISRCTN17754650. Prospectively registered on 03/08/2021 before recruitment began. https://doi.org/10.1186/ISRCTN17754650.

Protocol: Published: Yiend, J., Taher, R., Fialho, C., Hampshire, C., Hsu, C. W., Kabir, T., Keppens, J., McGuire, P., Mouchlianitis, E., Peters, E., Ricci, T., Shergill, S., Stahl, D., Vamvakas, G., Jacobsen, P., MPIT, & Avegen (2024). Assessing the efficacy and safety of STOP (successful treatment for paranoia)-an app-based cognitive bias modification therapy for paranoia: a randomised clinical trial protocol. *Trials*, *25*(1), 806. https://doi.org/10.1186/s13063-024-08570-3.

Data availability: *https://reshare.ukdataservice.ac.uk/857941/*.

Funding: This research was funded by the Medical Research Council (MRC) Biomedical Catalyst: Developmental Pathway Funding Scheme (DPFS), MRC Reference: MR/V027484/1. The funder provided only financial support for the conduct of the research and did not have a role in the collection, management, analysis or interpretation of data or in the writing of the manuscript.

Conflicts of Interests: The digital intervention evaluated in this study was developed by Avegen Health Ltd., who received funding for its development. The authors of this paper are independent from both the developers and the funders.

Acknowledgements*:* We want to thank the National Institute for Health and Care Research (NIHR) Biomedical Research Centre hosted at South London and Maudsley NHS Foundation Trust in partnership with King's College London. The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR, Department of Health and Social Care, the ESRC, or King's College London. This research was funded by the Medical Research Council (MRC) Biomedical Catalyst: Developmental Pathway Funding Scheme (DPFS), MRC Reference: MR/V027484/1. We would like to thank Avegen Health Ltd., the digital healthcare solutions company who created the trial version of the STOP App and hosted it on their proprietary platform, HealthMachine™, for the study duration. The Avegen group included, in alphabetical order: Nandita Kurup, Neeraj Apte, Neha Gupta. We acknowledge with thanks the contributions made to this work by the STOP study Lived Experience Advisory Panel (LEAP) supported by the McPin Foundation. The LEAP included, in alphabetical order: Al Richards, Alex Kenny, Edmund Brooks, Emily Curtis, and Vanessa Pinfold. Finally, we would also like to thank Elias Mouchlianitis and Ricardo Twumasi for their respective contributions.

References

1. Rus-Calafell M, Schneider S. Are we there yet?!—a literature review of recent digital technology advances for the treatment of early psychosis. *Mhealth*. 2020;6:3-3. doi:10.21037/mhealth.2019.09.14
2. Smith KA, Hardy A, Vinnikova A, et al. Digital Mental Health for Schizophrenia and Other Severe Mental Illnesses: An International Consensus on Current Challenges and Potential Solutions. *JMIR Ment Health*. 2024;11:e57155. doi:10.2196/57155

3. Clarke S, Hanna D, Mulholland C, Shannon C, Urquhart C. A systematic review and meta-analysis of digital health technologies effects on psychotic symptoms in adults with psychosis. *Psychosis*. 2019;11(4):362-373. doi:10.1080/17522439.2019.1632376
4. Craig TK, Rus-Calafell M, Ward T, et al. AVATAR therapy for auditory verbal hallucinations in people with psychosis: a single-blind, randomised controlled trial. *Lancet Psychiatry*. 2018;5(1):31-40. doi:10.1016/S2215-0366(17)30427-3
5. Garety PA, Edwards CJ, Jafari H, et al. Digital AVATAR therapy for distressing voices in psychosis: the phase 2/3 AVATAR2 trial. *Nat Med*. Published online October 28, 2024. doi:10.1038/s41591-024-03252-8
6. Garety P, Ward T, Emsley R, et al. Effects of SlowMo, a Blended Digital Therapy Targeting Reasoning, on Paranoia Among People With Psychosis. *JAMA Psychiatry*. 2021;78(7):714. doi:10.1001/jamapsychiatry.2021.0326
7. Greenwood KE, Gurnani M, Ward T, et al. The service user experience of SlowMo therapy: A co-produced thematic analysis of service users' subjective experience. *Psychology and Psychotherapy: Theory, Research and Practice*. 2022;95(3):680-700. doi:10.1111/papt.12393
8. Bucci S, Berry N, Ainsworth J, et al. Effects of Actissist, a digital health intervention for early psychosis: A randomized clinical trial. *Psychiatry Res*. 2024;339:116025. doi:10.1016/j.psychres.2024.116025
9. Bucci S, Barrowclough C, Ainsworth J, et al. Actissist: Proof-of-concept trial of a theory-driven digital intervention for psychosis. *Schizophr Bull*. 2018;44(5):1070-1080. doi:10.1093/schbul/sby032
10. Freeman D, Garety PA, Bebbington PE, et al. Psychological investigation of the structure of paranoia in a non-clinical population. *British Journal of Psychiatry*. 2005;186(5):427-435. doi:10.1192/bjp.186.5.427
11. Bentall RP, Rowse G, Shryane N, et al. The Cognitive and Affective Structure of Paranoid Delusions. *Arch Gen Psychiatry*. 2009;66(3):236. doi:10.1001/archgenpsychiatry.2009.1
12. Birchwood M, Trower P. The future of cognitive-behavioural therapy for psychosis: not a quasi-neuroleptic. *British Journal of Psychiatry*. 2006;188(2):107-108. doi:10.1192/bjp.bp.105.014985
13. GARETY PA, KUIPERS E, FOWLER D, FREEMAN D, BEBBINGTON PE. A cognitive model of the positive symptoms of psychosis. *Psychol Med*. 2001;31(2):189-195. doi:10.1017/S0033291701003312
14. Savulich G, Freeman D, Shergill S, Yiend J. Interpretation Biases in Paranoia. *Behav Ther*. 2015;46(1):110-124. doi:10.1016/j.beth.2014.08.002
15. Johnson J. The Validity of Major Depression With Psychotic Features Based on a Community Study. *Arch Gen Psychiatry*. 1991;48(12):1075. doi:10.1001/archpsyc.1991.01810360039006
16. SHER L. Manic-Depressive Illness: Bipolar Disorders and Recurrent Depression, Second Editionby Frederick K. Goodwin, M.D., and Kay Redfield Jamison, Ph.D. New York,

Oxford University Press, 2007. *American Journal of Psychiatry*. 2008;165(4):541-542. doi:10.1176/appi.ajp.2007.07121846
17. Hamner MB, Frueh BC, Ulmer HG, Arana GW. Psychotic features and illness severity in combat veterans with chronic posttraumatic stress disorder. *Biol Psychiatry*. 1999;45(7):846-852. doi:10.1016/S0006-3223(98)00301-1
18. van Os J, Verdoux H, Maurice-Tison S, et al. Self-reported psychosis-like symptoms and the continuum of psychosis. *Soc Psychiatry Psychiatr Epidemiol*. 1999;34(9):459-463. doi:10.1007/s001270050220
19. Appelbaum PS, Robbins PC, Roth LH. Dimensional Approach to Delusions: Comparison Across Types and Diagnoses. *American Journal of Psychiatry*. 1999;156(12):1938-1943. doi:10.1176/ajp.156.12.1938
20. Freeman D, Garety PA, Kuipers E, Fowler D, Bebbington PE. A cognitive model of persecutory delusions. *British Journal of Clinical Psychology*. 2002;41(4):331-347. doi:10.1348/014466502760387461
21. Wessely S, Buchanan A, Reed A, et al. Acting on Delusions. I: Prevalence. *British Journal of Psychiatry*. 1993;163(1):69-76. doi:10.1192/bjp.163.1.69
22. Castle DJ, Phelan M, Wessely S, Murray RM. Which Patients with Non-affective Functional Psychosis are not Admitted at First Psychiatric Contact? *British Journal of Psychiatry*. 1994;165(1):101-106. doi:10.1192/bjp.165.1.101
23. Trotta A, Kang J, Stahl D, Yiend J. Interpretation Bias in Paranoia: A Systematic Review and Meta-Analysis. *Clinical Psychological Science*. 2021;9(1):3-23. doi:10.1177/2167702620951552
24. van der Gaag M, Valmaggia LR, Smit F. The effects of individually tailored formulation-based cognitive behavioural therapy in auditory hallucinations and delusions: A meta-analysis. *Schizophr Res*. 2014;156(1):30-37. doi:10.1016/j.schres.2014.03.016
25. Mehl S, Werner D, Lincoln TM. Does Cognitive Behavior Therapy for psychosis (CBTp) show a sustainable effect on delusions? A meta-analysis. *Front Psychol*. 2015;6. doi:10.3389/fpsyg.2015.01450
26. Burgess-Barr S, Nicholas E, Venus B, et al. International rates of receipt of psychological therapy for psychosis and schizophrenia: systematic review and meta-analysis. *Int J Ment Health Syst*. 2023;17(1):8. doi:10.1186/s13033-023-00576-9
27. Turner DT, van der Gaag M, Karyotaki E, Cuijpers P. Psychological Interventions for Psychosis: A Meta-Analysis of Comparative Outcome Studies. *American Journal of Psychiatry*. 2014;171(5):523-538. doi:10.1176/appi.ajp.2013.13081159
28. Elkis H. Treatment-Resistant Schizophrenia. *Psychiatric Clinics of North America*. 2007;30(3):511-533. doi:10.1016/j.psc.2007.04.001
29. Moritz S, Woodward TS. Metacognitive training in schizophrenia: from basic research to knowledge translation and intervention. *Curr Opin Psychiatry*. 2007;20(6):619-625. doi:10.1097/YCO.0b013e3282f0b8ed

30. Waller H, Freeman D, Jolley S, Dunn G, Garety P. Targeting reasoning biases in delusions: A pilot study of the Maudsley Review Training Programme for individuals with persistent, high conviction delusions. *J Behav Ther Exp Psychiatry*. 2011;42(3):414-421. doi:10.1016/j.jbtep.2011.03.001
31. Hallion LS, Ruscio AM. A meta-analysis of the effect of cognitive bias modification on anxiety and depression. *Psychol Bull*. 2011;137(6):940-958. doi:10.1037/a0024355
32. Yiend J, Lam CLM, Schmidt N, et al. Cognitive bias modification for paranoia (CBM-pa): a randomised controlled feasibility study in patients with distressing paranoid beliefs. *Psychol Med*. 2023;53(10):4614-4626. doi:10.1017/S0033291722001520
33. Yiend J, Trotta A, Meek C, et al. Cognitive Bias Modification for paranoia (CBM-pa): study protocol for a randomised controlled trial. *Trials*. 2017;18(1):298. doi:10.1186/s13063-017-2037-x
34. Hsu et al. User-centered Development of STOP (Successful Treatment for Paranoia): Material Development and Usability Testing for a Digital Therapeutic for Paranoia. *JMIR Human Factors* 2023;10:e45453. https://humanfactors.jmir.org/2023/1/e45453
35. Fenigstein A, Vanable PA. Paranoia and self-consciousness. *J Pers Soc Psychol*. 1992;62(1):129-138. doi:10.1037/0022-3514.62.1.129
36. Taher R, Hall CL, Bergin ADG, et al. Developing a process for assessing the safety of a digital mental health intervention and gaining regulatory approval: a case study and academic's guide. *Trials*. 2024;25(1):604. doi:10.1186/s13063-024-08421-1
37. Yiend J, Trotta A, Meek C, et al. Cognitive Bias Modification for paranoia (CBM-pa): study protocol for a randomised controlled trial. *Trials*. 2017;18(1):298. doi:10.1186/s13063-017-2037-x
38. Yiend J, TR, FC, HC, HCW, KT, KJ, MP, ME, PE, RT, SS, SD, JP, VG, PM, & GA (Accepted/In press). Assessing the efficacy and safety of STOP (Successful Treatment for Paranoia) - an app-based cognitive bias modification therapy for paranoia: a randomized clinical trial protocol. *Trials*.
39. Eysenbach G. CONSORT-EHEALTH: Improving and Standardizing Evaluation Reports of Web-based and Mobile Health Interventions. *J Med Internet Res*. 2011;13(4):e126. doi:10.2196/jmir.1923
40. Hopewell S, Chan AW, Collins GS, et al. CONSORT 2025 statement: updated guideline for reporting randomised trials. *BMJ*. 2025;389:e081123. doi:10.1136/bmj-2024-081123
41. Kay SR, Fiszbein A, Opler LA. The Positive and Negative Syndrome Scale (PANSS) for Schizophrenia. *Schizophr Bull*. 1987;13(2):261-276. doi:10.1093/schbul/13.2.261
42. Mathews A, Mackintosh B. Induced emotional interpretation bias and anxiety. *J Abnorm Psychol*. 2000;109(4):602-615.
43. Eysenck MW, Mogg K, May J, Richards A, Mathews A. Bias in interpretation of ambiguous sentences related to threat in anxiety. *J Abnorm Psychol*. 1991;100(2):144-150. doi:10.1037/0021-843X.100.2.144

44. Struman E. The capacity to consent to treatment and research: A review of standardized assessment tools. *Clin Psychol Rev*. 2005;25(7):954-974. doi:10.1016/j.cpr.2005.04.010
45. Bush B, Shaw S, Cleary P, Delbanco TL, Aronson MD. Screening for alcohol abuse using the cage questionnaire. *Am J Med*. 1987;82(2):231-235. doi:10.1016/0002-9343(87)90061-1
46. Imran JB, Richmond RE, Madni TD, et al. Determining suicide risk in trauma patients using a universal screening program. *Journal of Trauma and Acute Care Surgery*. 2018;85(1):182-186. doi:10.1097/TA.0000000000001899
47. Hettige NC, Bani-Fatemi A, Kennedy JL, De Luca V. Assessing the risk for suicide in schizophrenia according to migration, ethnicity and geographical ancestry. *BMC Psychiatry*. 2017;17(1):63. doi:10.1186/s12888-016-1180-3
48. R. I. Kotov SBB& DBW. *Multidimensional Iowa Suggestibility Scale (MISS) Brief Manual*.; 2004.
49. Devilly GJ, Borkovec TD. Psychometric properties of the credibility/expectancy questionnaire. *J Behav Ther Exp Psychiatry*. 2000;31(2):73-86. doi:10.1016/S0005-7916(00)00012-4
50. Rude SS, Valdez CR, Odom S, Ebrahimi A. Negative cognitive biases predict subsequent depression. Cognitive Therapy and Research. *Cognit Ther Res*. 2003;27(4):415-429. doi:10.1023/A:1025472413805
51. Rude SS, Wenzlaff RM, Gibbs B, Vane J, Whitney T. Negative processing biases predict subsequent depressive symptoms. *Cogn Emot*. 2002;16(3):423-440. doi:10.1080/02699930143000554
52. Freeman D, Loe BS, Kingdon D, et al. The revised Green *et al.* , Paranoid Thoughts Scale (R-GPTS): psychometric properties, severity ranges, and clinical cut-offs. *Psychol Med*. 2021;51(2):244-253. doi:10.1017/S0033291719003155
53. Zigmond AS, Snaith RP. The Hospital Anxiety and Depression Scale. *Acta Psychiatr Scand*. 1983;67(6):361-370. doi:10.1111/j.1600-0447.1983.tb09716.x
54. Freeman D, Bird JC, Loe BS, et al. The Dunn Worry Questionnaire and the Paranoia Worries Questionnaire: new assessments of worry. *Psychol Med*. 2020;50(5):771-780. doi:10.1017/S0033291719000588
55. Tennant R, Hiller L, Fishwick R, et al. The Warwick-Edinburgh Mental Well-being Scale (WEMWBS): development and UK validation. *Health Qual Life Outcomes*. 2007;5(1):63. doi:10.1186/1477-7525-5-63
56. Law H, Neil ST, Dunn G, Morrison AP. Psychometric properties of the Questionnaire about the Process of Recovery (QPR). *Schizophr Res*. 2014;156(2-3):184-189. doi:10.1016/j.schres.2014.04.011
57. Keetharuth AD, Brazier J, Connell J, et al. Recovering Quality of Life (ReQoL): a new generic self-reported outcome measure for use with people experiencing mental health difficulties. *The British Journal of Psychiatry*. 2018;212(1):42-49. doi:10.1192/bjp.2017.10

58. Ben-Zeev D, Brenner CJ, Begale M, Duffecy J, Mohr DC, Mueser KT. Feasibility, Acceptability, and Preliminary Efficacy of a Smartphone Intervention for Schizophrenia. *Schizophr Bull*. 2014;40(6):1244-1253. doi:10.1093/schbul/sbu033
59. Mallinckrodt CH, Clark WS, David SR. ACCOUNTING FOR DROPOUT BIAS USING MIXED-EFFECTS MODELS. *J Biopharm Stat*. 2001;11(1-2):9-21. doi:10.1081/BIP-100104194
60. Savulich G, Shergill SS, Yiend J. Interpretation Biases in Clinical Paranoia. *Clinical Psychological Science*. 2017;5(6):985-1000. doi:10.1177/2167702617718180

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- STOPmainoutcomeSupplementary.docx
- 303876STOPStudyProtocolV3.1290622revisedCLEAN.pdf
- CI20220033GBMHRAOutcomeNoObjection.pdf
- 303876STOPStudyProtocolV1.0061121.pdf