Statistical Analysis Plan

Study Title: The effects of PF-04995274 on emotional processing in treatment-resistant, medicated, depressed patients (RESTART study).

CT REGISTRATION: NCT03515733. ETHICS REF: 18/SC/0074

Objectives

The primary aim of the study is to investigate the effects of 7 days of PF-04995274 administration (adjunctive to SSRI/SNRI medication) on behavioural measures of non-emotional and emotional cognition, specifically memory performance on an auditory verbal learning task and performance (including accuracy and reaction times) on a facial expression recognition task. The secondary aim of the study is to investigate the effects of 7 days of PF-04995274 administration (adjunctive to SSRI/SNRI medication) on other behavioural measures of non-emotional and emotional cognition.

Brief summary of design

This study uses a double-blind, placebo-controlled, randomised between-groups design. Participants are patients who fulfil criteria for current episode of Major Depressive Disorder (MDD) and are currently taking an SSRI/SNRI and have failed to response clinically. Participants will be randomised to receive 7-9 days treatment with either PF-04995274 (15 mg daily) or a matched placebo. This study includes three visits in total: (a) Screening Visit; (b) First Dose Visit; (c) Research Visit Two. All visits will take place at the Warneford Hospital, Oxford University Department of Psychiatry.

Determination of Sample Size

We will recruit 50 participants to the study (25 on PF-0499574 and 25 on placebo). Participants who withdraw during the study or do not provide complete data-sets will be replaced. Based on data acquired in Harmer et al., (2004) comparing citalopram to placebo, if we aim for 0.9 power and a 0.05 false positive rate, a suggested group sample size is 19 (G*power) to ensure determination of group level differences at this variable if they exist. As 5HT4 agonism is less well studied, and to account for the exclusion of low quality data before analysis, we will aim for 25 individuals with complete datasets per group (total sample size of 50).

Data Cleaning

- Will be performed prior to unblinding
- Outliers will be excluded on a per task basis
- For all behavioural data, excluding the EPS, cut-off thresholds will be determined based on a visual inspection of a histogram plot, examining thresholds for:
  - *Trials with unusually low response times*
  - *Trials with unusually high response times*
  - *Proportion of missing/removed trials per participant*
  - *Abnormally low mean accuracy (or equivalent outcome) per participant*
  - *Abnormally high mean reaction time per participant*
- For all self-report data, extreme outliers indicating invalid data entry will be determined based on a visual inspection of a histogram plot
- For emotion potentiated startle data, two researchers will independently a) distinguish startle blink response from noise and decide whether a response could have been seen, had one occurred, excluding trials if no response could not be seen and b) determine if there is a blink response or if the trial should be recorded as a non-response. If there is disagreement, a third researcher will make a final decision.

Behavioural Analysis

Below is a non-exhaustive list of outcomes and analyses which will be conducted.

Behavioural Task	**Outcomes**	**Analysis** • All key endpoints will be summarized (mean, standard deviation) in tables and bar charts (mean ±SEM) • Conducted in R (version will be confirmed in publications)
Facial Expression Recognition Task (FERT) Recognition of computer-based positive and negative facial expressions P1vital® Limited Products	**Unbiased hit rate**, as described by Wagner (1993) – a measure of emotion identification accuracy which accounts for response bias i.e. any general tendency to identify the emotion when it is not present. Calculated as proportion of correct hits * (number of hits/all hits and misses), for each facial expression category. **% correct and response bias** will also be reported individually.	Repeated measures analyses of variance (ANOVAs): • Between-subject factor – 2 levels: Treatment group (PF-04995274 or placebo) • Within-subject factor – 7 levels (Fear, anger, happy, surprise, disgust, sad, neutral)
	Misclassifications: Number of responses to each facial expression category incorrectly classified as another facial expression category e.g. identifying a fearful face as surprised	
	Reaction time (ms) for trials with correct responses.	
Auditory Verbal Learning Task (AVLT) Recall of words read aloud	**Number of words recalled -** List A immediate recall trials	Repeated measures analyses of variance (ANOVAs): • Between-subject factor – 2 levels: Treatment group (PF-04995274 or placebo) • Within-subject factor - 5 levels (List A immediate recall trials 1-5)

Pen and paper	**Number of words recalled -** List A short delay **Number of words recalled -** List A long delay	Repeated measures analyses of variance (ANOVAs): • Between-subject factor – 2 levels: Treatment group (PF-04995274 or placebo) • Within-subject factor - 2 levels (List A short and long delay trials)
	Number of intrusions (words incorrectly recalled) across List A acquisition trials	Independent samples t-tests
	Number of repetitions (words repeated within the same trial) across List A acquisition trials	
	Number of words recalled - List B recall	
	Number of hits and false alarms in the delayed recognition test	
Probabilistic Instrumental Learning Task (PILT) Reward sensitivity Neurobehavioral Systems Presentation software (https://www.neurobs.com)	**% Accuracy** (correct or incorrect symbol choice) Correct = symbol associated with high probability of winning or low probability of losing	Independent samples t-tests
	Proportion of group choosing correct symbol per trial	The proportion will be calculated, and plotted on a learning curve to determine where learning plateaus. Repeated measures analyses of variance (ANOVAs) - trials where learning has plateaued • Between-subject factor – 2 levels: Treatment group (PF-04995274 or placebo) • Within-subject factor – 2 levels: Condition (win or loss)
	Learning rate from reinforcement learning model	Repeated measures analyses of variance (ANOVAs) • Between-subject factor – 2 levels: Treatment group (PF-04995274 or placebo) • Within-subject factor – 2 levels: Condition (win or loss)

	Decision temperature parameters from reinforcement learning model	Repeated measures analyses of variance (ANOVAs) • Between-subject factor – 2 levels: Treatment group (PF-04995274 or placebo) • Within-subject factor – 2 levels: Condition (win or loss)
	Amount won	Independent samples t-tests
	Amount lost	
	Total monetary amount earned	
Emotional Categorisation Task (ECAT) Categorisation of emotional words P1vital® Limited Products	**% Accuracy** – words correctly identified as positive or negative	Mixed model analyses of variance (ANOVAs). • Between-subject factor – 2 levels: Treatment group (PF-04995274 or placebo) • Within-subject factor – 2 levels: Word valence (positive or negative)
	Reaction time	
Emotional Recall Task (EREC) Recall of emotional words from ECAT P1vital® Limited Products	**Number of hits** (words recalled correctly)	
	Number of false alarms (words recalled incorrectly)	
Emotional Recognition Task (EMEM) Recognition of emotional words from ECAT P1vital® Limited Products	**Number of hits** (words recognised correctly)	
	Number of false alarms (words recognised incorrectly)	
	Reaction time	

Facial Dot Probe Task (FDOT) Vigilance to fearful and happy faces P1vital® Limited Products	**Vigilance scores** derived from reaction time – e.g. bias scores calculated by subtracting median reaction times in congruent trials (i.e. the probe appears behind the emotional expression) from those in incongruent trials (i.e. the probe appears behind the neutral expression)	Mixed models analyses of variance (ANOVAs). • Between-subject factor – 2 levels: Treatment group (PF-04995274 or placebo) • Within-subject factors – 2 levels: Emotion (positive or negative) – 2 levels: Probe duration (masked or unmasked)
Emotion Potentiated Startle (EPS) EMG data, in response to white noise during positive or negative images San Diego Instruments, San Diego, CA, USA	**Raw amplitude** of startle response **Z-transformed amplitude** of startle response **Latency** of startle response **(ms)**	Mixed models analyses of variance (ANOVAs). • Between-subject factor – 2 levels: Treatment group (PF-04995274 or placebo) • Within-subject factors – 2 levels: Trial type (positive, negative, neutral)
Oxford Memory Test (OMT) Visual short term spatial memory Oxford Memory Test application "Short_Fractals1" – modified from "What was where task" (Pertzov et al., 2013) running on iOS 12.3.1	**Proportion of correct probe selections** **Absolute error** for probe location **Reaction Time**	Mixed models analyses of variance (ANOVAs). • Between-subject factor – 2 levels: Treatment group (PF-04995274 or placebo) • Within-subject factors – 2 levels: Trial condition (1 or 3 memory probes)

Self-report or researcher-observed scale – all completed on Qualtrics.XM (https://www.qualtrics.com) except HAM-D		
Eysenck Personality Questionnaire (EPQ)	Total score for each dimension	Report descriptives for each group
State and Trait Anxiety Inventory – Trait subscale (STAI-T)	Total score	
Becks Depression Inventory (BDI)	Total score	
Hamilton Depression Scale (HAM-D) Pen and paper – scored by research team	Total score	
Snaith-Hamilton Pleasure Scale (SHAPS)	Total score	
State and Trait Anxiety Inventory – State subscale (STAI-S)	Total score	Mixed model ANOVAs: • Between-subject factor – 2 levels: Treatment group (PF-04995274 or placebo) • Within-subject factors – 4 levels: Time condition (Pre-scan, Post-scan, Pre-ETB, Post-ETB)
Positive and Negative Affect Scale (PANAS)	Total score for positive and negative subscales	
Visual Analogue Scales (VAS)	Total score for each VAS (happy, sad, hostile, alert, anxious, calm)	
Side effects	Presence of side effect Severity of side effect Belief in relationship to drug	Descriptive report of frequency of side-effects for each group at four time points (baseline, pre-dose, post-dose and all other study days combined). A generalised linear model will be used to analyse side effects, with presence of side-effect as outcome and predictors including treatment group (PF-04995274 or placebo) and time point (baseline, pre-dose, post-dose, day 2/3/4/5/6/7/8/9). For side effects significantly associated with group and time-point, we will investigate severity and belief in relationship to study drug.

When conducting ANOVAs, the Greenhouse-Geisser procedure will be used to correct the degrees of freedom where assumptions of equality of variance are violated. If there is a significant group x condition interaction found in ANOVAs. Post hoc independent samples t tests will be performed to follow up any significant interactions. We will not use the Bonferroni correction for multiple comparisons for post-hoc tests. When conducting t tests, degrees of freedom will be corrected where the assumption of equal variances between groups is violated (i.e. Levene's Test is significant).

Record of version changes and unblinding

Date	Version	Blinding Status	Comments
28th October 2022	1.0	Team blinded, barring unblinding for study medics (AdeC, PC and BG) where necessary.	Data collection complete. First version of complete stats plans. Uploaded to clinicaltrials.gov and OSF.

Study Team involved in analysis

AdeC - Dr Angharad de Cates – DPhil Student, Study Medic

AG - Dr Amy Gillespie – Post-doctoral Researcher

BG - Dr Beata Godlewska - Study Medic

CH - Professor Catherine Harmer - Principal Investigator

MB - Merethe Blandhol – Research Assistant

PC - Professor Phil Cowen – Principal Investigator, Study Medic

SM - Dr Susannah Murphy – Senior Research Fellow