Statistical Analysis Plan

Study Title: The effects of PF-04995274 on emotional processing in un-medicated depressed patients (RESTAND study).

CT REGISTRATION: NCT3516604. ETHICS REF: 18/SC/0076

Objectives

The primary aim of the study is to investigate the effects of 7 days of PF-04995274 administration on neural activity related to non-emotional and emotional cognition, specifically during a novel vs familiar fMRI task and an emotional faces fMRI task. The secondary aim is to investigate the effects of 7 days of PF-04995274 administration 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 computer-based tasks of emotional processing.

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 unmedicated. Participants will be randomised to receive 7-9 days treatment with either PF-04995274 (15 mg daily), citalopram (20mg) or a matched placebo. This study includes four visits in total: (a) Screening Visit; (b) First Dose Visit; (c) Research Visit One, and (d) Research Visit Two. All visits will take place at the Warneford Hospital, Oxford University Department of Psychiatry.

Determination of Sample Size

We will recruit 75 participants to the study (25 on PF-0499574, 25 on citalopram 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 75).

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.

Imaging Analysis

Faces Task **Neural response to emotional faces**	**Memory Encoding Task** **Neural response to novel vs. familiar scenes**	**Resting state connectivity**	**Relative and global cerebral blood flow**
Blood Oxygen Level Dependent (BOLD) signal in areas including the amygdala, anterior cingulate cortex, and orbitofrontal cortex	Blood Oxygen Level Dependent (BOLD) signal to scenes that have previously been seen compared to novel scenes in areas including the hippocampus and parahippocampal regions	Resting state connectivity within and between networks, identified via correlations between spontaneous BOLD activity in spatially independent regions while participants are not actively engaged in an experimental task	Relative and global cerebral blood flow identified using Arterial Spin Labelling (ASL)
Data collection			
3 Tesla Siemens Prisma Sequence acquisition parameters stored on OSF.			
MRI stimulus presentation: *Psychopy ((http://www.psychopy.org)*		NA	NA
Ancillary physiological data will be collected during MRI scanning and may be used to remove noise in the brain data. These data include pulse oximetry, respiratory activity and eye gaze location. MRI physiological measurement acquisition: *Biopac MP150 system, software level (https://www.biopac.com), Eyelink ® (https://www.sr-research.com)*			
Data preparation and pre-processing			
• Data transfer from the scanner server to the high performance computing analysis cluster in Digital Imaging and Communications in Medicine (DICOM) format. • Visual checks of fMRI data - excessive movement, other artefacts such as signal drop out, all data files present. • Data converted to Brain Imaging Data Structure (BIDS) format nifti files using heudiconv.*(https://github.com/nipy/heudiconv)*			

<table>
<tr><td colspan="3">

- Formal MRI quality assessment - *MRIQC package (https://mriqc.readthedocs.io/en/stable/index.html*), with data considered for rejection if falls outside the normal range of values in the derived image quality metrics.
- Final data inclusion meeting with Principal Investigator prior to unblinding to confirm exclusion of data from specific tasks on a per task basis—that is, a participant may have data excluded for one task, but have all other data available for analysis.

</td></tr>
<tr><td>

Pre-processing will be carried out using the FSL pipeline for task data

This will include:

- Brain extraction
- Spatial smoothing (4-5 mm full width half maximum (FWHM) kernel which is appropriate to detect effects in small structures such as the amygdala)
- High pass temporal filtering (calculated by FSL to define the highest frequency which would be expected in the data given the temporal structure of the task - calculated using the 'cutoffcalc' algorithm to retain 90% of the expected signal after filtering)
- B0 unwarping using fieldmap images
- Motion correction MCFLIRT
- Registration of EPI images to T1w images and to MNI standard space

FSL (https://process.innovation.ox.ac.uk/software/p/9564/fslv5/1)

</td><td>

We will use fmiprep if necessary for preprocessing with similar steps to task data.

Tedana will be used to produce a denoised optimally combined file for each participant

</td><td>

We will use fmiprep if necessary for preprocessing with similar steps to task data.

</td></tr>
<tr><td colspan="3">***Modelling – first and second level analysis***</td></tr>
<tr><td>

Task specific regressors, describing the onset and duration of task relevant events will be defined for each participant.

These regressors will be convolved with a standard haemodynamic response function, filtered using the same highpass filter applied to the functional data, and regressed against the preprocessed functional data.

First level model may include nuisance regressors representing the sources of physiological noise in the fMRI signal. These regressors would be created using the PNM tool of the FSL package from pulse oximetry and respiratory bellows data collected during the tasks, and

</td><td>

Preprocessed cleaned data will then be temporally concatenated across subjects and decomposed into independent components (ICs) using FSL Melodic. IC maps will be identified as being analogous to major resting state networks.

</td><td>

BASIL tools from the FSL suite will be used to create a statistical map estimate of local cerebral blood flow for each participant.

BASIL (latest version available)

</td></tr>
</table>

<table>
<tr><td colspan="2">custom scripts eye gaze direction. Gaze information may be used to remove trials where participants were not correctly fixating on the task, or account for variance in activation between participants or groups where systematic variations in gaze location are observed. The purpose of these regressors is to account for noise in the data and thus increase the sensitivity of the analyses.

The output of the first level analyses will be contrast maps (i.e. contrasting the effects of the regressors in the model), one for each participant.

The model is regressed against the smoothed, pre-processed BOLD data to generate maps (betas) of how well the data in each voxel is explained by the model (after the effects of the nuisance regressors are removed if these are being used). Explicit task effects are determined by the beta weights of each individual task relevant event, which are contrasted between conditions to address the experimental questions. Testing of the GLM will employ FMRIB's Improved Linear Model (FILM) prewhitening of the BOLD data to allow robust estimation of task relevant betas with consideration of autocorrelation in the voxel time series.</td><td rowspan="2"></td><td rowspan="2">(https://fsl.fmrib.ox.ac.uk/fsl/fslwiki/BASIL);</td></tr>
<tr><td>Two regressors coding for (1) blocks of fearful faces and (2) blocks of happy faces.

Contrasts: i) all faces (the mean activation); ii) fear only; iii) happy only; iv) fear > happy; v) fear < happy; vi) fear < baseline; vii) happy < baseline.</td><td>Two regressors coding for (1) blocks of novel scenes and (2) blocks of repeated ("familiar") scenes.

Contrasts: i) all images (the mean activation); ii) novel only; iii) familiar only; iv) novel > familiar; v) novel < familiar; vi) familiar < baseline ; vii) novel < baseline.</td></tr>
<tr><td colspan="2">Inputs: first level contrast maps of model fit

The following second level analyses will be run:

a) whole brain analyses (i.e. looking at all voxels in the brain);</td><td>Focus on networks which have been identified as showing significant differences between patients with depression and healthy controls, and</td><td>Inputs: first level maps of cerebral blow flow

Same analyses as task-based fMRI.</td></tr>
</table>

b) small volume corrected analyses (i.e. image based analyses limited to the prespecified regions of interest listed below; c) a region of interest analysis (i.e. using the mean activity within the prespecified regions of interest listed below). We will also consider adding grey matter and / or ASL maps and / or gender to the confounder variables as a confirmatory analysis. The following prespecified regions of interest will be used: • Left and right amygdala (as defined in the Harvard-Oxford Subcortical Atlas) – faces only • Hippocampus (defined in the Harvard-Oxford Subcortical Atlas) • Medial prefrontal cortex (defined in the Harvard-Oxford Subcortical Atlas) – faces only • Orbitofrontal cortex (defined in the Harvard-Oxford Subcortical Atlas) – faces only For functional masks, we will use a mask for each contrast of interest created by multiplying mean activation for all participants for this contrast by a Harvard-Oxford Histological atlas anatomical mask at a 50% threshold.	which were identified in our previous work with a 5HT4 agonist: • default mode network • salience network • cognitive control network We will conduct exploratory analysis of all commonly identified resting state networks as controls. Exploratory seed analysis may be conducted as appropriate.	
Summary statistics (estimates of activity - model beta parameter estimates) will be extracted and entered into a repeated measures ANOVA. Image based statistical analyses will be corrected for multiple comparisons within the region of brain analysed. For whole brain analyses this will be all voxels within the brain, for analyses limited to a prespecified anatomical location (i.e. "small volume corrected analyses") this will be across the voxels within the prespecified mask (prespecified regions of interest). Multiple comparison correction will be achieved using threshold free cluster enhancement (TFCE) or cluster-based analysis while controlling for family wise error rate with a z threshold of 3.1 and a corrected p value of < 0.05. If randomise is used, we will interrogate for group differences by non-parametric permutation testing (5000 permutations) with FSL randomise (TFCE, p<0.05 corrected).	We will test for statistically significant differences between the groups across all identified networks using FSL's randomize permutation-testing tool (5000 permutations). Threshold-Free-Cluster-Enhancement (TFCE) approach will be used and a family-wise-error corrected cluster significance threshold of $p < 0.05$ applied to the suprathreshold clusters to correct for multiple comparisons at the voxel level. The GLM will include the groups of interest comparison.	Same analyses as task-based fMRI.

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 – 3 levels: Treatment group (PF-04995274, citalopram 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 – 3 levels: Treatment group (PF-04995274, citalopram 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 – 3 levels: Treatment group (PF-04995274, citalopram 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 – 3 levels: Treatment group (PF-04995274, citalopram 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 – 3 levels: Treatment group (PF-04995274, citalopram 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 – 3 levels: Treatment group (PF-04995274, citalopram 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 – 3 levels: Treatment group (PF-04995274, citalopram 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 – 3 levels: Treatment group (PF-04995274, citalopram 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	Mixed models analyses of variance (ANOVAs). • Between-subject factor – 3 levels: Treatment group (PF-04995274, citalopram or placebo) • Within-subject factors – 2 levels: Trial type (positive, negative, neutral) At the time of writing, we know there were technical difficulties with the EMG machine during data collection leading to fewer datasets and reduced statistical power to identify an interaction. Therefore, we pre-specify looking at each group separately in repeated-measures t-tests: • PF-04995274 vs placebo • Citalopram vs placebo • PF-04995274 vs citalopram
	Z-transformed amplitude of startle response	
	Latency of startle response **(ms)**	
Oxford Memory Test (OMT)	**Proportion of correct probe selections**	Mixed models analyses of variance (ANOVAs). • Between-subject factor – 3 levels: Treatment group (PF-04995274, citalopram or placebo)
	Absolute error for probe location	

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	**Reaction Time**	• Within-subject factors – 2 levels: Trial condition (1 or 3 memory probes)
Behavioural data for scanner task - novel vs. familiar images Neurobehavioral Systems Presentation software (https://www.neurobs.com)	**Accuracy** - Percentage of images correctly recognized post-scanner	Independent samples t-tests

Self-report or researcher-observed scale – all completed on Qualtrics.XM (https://www.qualtrics.com) except HAM-D		
Edinburgh Handedness Inventory (EHI)	Total score	Report descriptives for each group
Eysenck Personality Questionnaire (EPQ)	Total score for each dimension	
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 – 3 levels: Treatment group (PF-04995274, citalopram 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, citalopram 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

IGS - Isabelle Goodall-Summers – FHS Student

MB - Merethe Blandhol – Research Assistant

PC - Professor Phil Cowen – Principal Investigator, Study Medic

SM - Dr Susannah Murphy – Senior Research Fellow