Effects of affective training during social interactions on momentary anxiety in youth at risk for social anxiety: a randomised controlled trial

Statistical Analysis Plan

Version 0.23

8 June 2026

1 Introduction

1.1 Purpose

This statistical analysis plan (SAP) contains details of the main pre-specified statistical analyses for the Social Surprises trial. This SAP describes the statistical analysis of the clinical outcomes; it does not contain details of any qualitative analyses. The SAP does not preclude the undertaking of further ad hoc or exploratory analyses, although the results of any such analyses should be interpreted with caution. Furthermore, the SAP does not preclude the adaptation of any part of the trial analysis should situations arise in which such adaptation is deemed necessary; any such adaptation will be transparent and fully justified.

1.2 Protocol version

Full details of the trial design, population, intervention, comparison and outcome variables may be found in the protocol.

1.3 Trial registration

This trial was prospectively registered with ISRTCN (XXX).

1.4 Authorship

This SAP has been written by Argyris Stringaris (AS) with input/oversight from Gareth Ambler (GA) following the guidelines of Gamble *et al* (JAMA 2017. doi:10.1001/jama.2017.18556).

1.5 SAP Revisions

Version	Date	Changes
1.0	XXX	

1.6 Signatures

Authorised by:	Signature	Date
Prof A Stringaris Chief Investigator		
Prof. Gareth Ambler Lead Statistician		
XXX		
XXXX		

2 Trial Summary

2.1 Title

Effects of affective training during social interactions on momentary anxiety in youth at risk for social anxiety: a randomised controlled trial (short title: Social Training of Affective Response (STAR) Trial).

2.2 Aims

The specific objectives of the STAR Trial are to:

- Assess the efficacy of social surprises intervention to improve momentary anxiety at outcome (two main blocks of the intervention)
- Assess the efficacy of social surprises intervention to improve momentary mood at outcome (two main blocks of the intervention)
- Assess the efficacy of social surprises intervention to provide affective resilience to negative feedback within the task (at the end of the negative section of the third, 'challenge' block).
- Assess the efficacy of social surprises intervention to improve momentary mood and anxiety at the end of the third block, following a sequence of negative feedback.

2.3 Population

Young people (18-25) at high risk of social anxiety

2.3.1 Inclusion criteria

1. Elevated symptoms of Social Anxiety Disorder as assessed by the age-appropriate Social Phobia Inventory (SPIN), defined as scoring above 19, which is the suggested cut-off for social anxiety.

2. Age between 18 and 25 (inclusive)

A) Confirmed by the participant during screening

3. Regular access to a computer connected to the Internet and a mobile phone to receive SMS messages.

A) Confirmed by the participant during screening

4. Residence/Current location in the UK or US

5. For online recruitment sites like prolific we will:

a) include a prolific approval rate of 95% as an inclusion criterion. This refers to the percentage participant's previous prolific jobs that have been approved.
b) require participants to have previously completed at least 10 jobs on prolific.
c) include 2 attention checks. If participants fail both, they are excluded from the study.

2.3.2 Exclusion criteria

Individuals are not eligible if they:
1. Don't speak sufficient English to complete questionnaires and engage in the English language-based task

A) Confirmed by the participant during screening

2. Have sensory or other deficits that preclude them from processing the computerized tasks

A) Confirmed by the participant during screening

3 Study Methods

3.1 Design

A double blind, sham-controlled study of a computerised task that generates social surprises in a virtual social interaction (active group) compared to a task that is identical except for the generation of social surprises (sham)of individuals recruited via: a) local communities (aged 18-25: tested online or in person), b) via online testing platforms such as the Amazon's Mechanical Turk (MTurk), Testable Minds, CloudResearch, or Prolific (aged 18-25; tested online). These are online platforms commonly used for these research purposes. Participants from online platforms will be screened online prior to the testing session.

3.1.1 Intervention

The intervention is inspired by the success of cognitive therapy for social anxiety, leading to a reduction of both anxiety and depressive symptoms in those patients.

The intervention is predicated on the idea that positive social surprises are the mechanism that leads to this improvement: positive social surprises are outcomes in a social situation that are better than one expects. Our intervention seeks to emulate such social situations to test this hypothesis and its usefulness for patients.

Participants will be sat in front of a computer and will interact with three virtual players across three blocks during the task. On each trial, they will first rate their expectation of how positively they believe the virtual player will rate them. Next they will respond to a self-referential question while being video recorded for 15 seconds. Afterward, they will receive feedback from the virtual player and then report their current mood and anxiety. Importantly, the feedback will gradually change across blocks: for the two 'main' blocks the feedback given will improve throughout the block, mirroring a common real-world social dynamic, whereby social reward increases the longer one remains in an interaction (e.g. meeting a stranger who initially feels neutral but becomes progressively more engaged over time). In a third, 'challenge' block, participants will interact with a third virtual player and receive a sequence of negative feedback. This block serves the purpose of testing affective resilience to negative feedback following the main blocks of the intervention.

3.1.2 Comparison (treatment as usual)

The active control arm will be identical in all respects to the intervention except the feedback in the two 'main' blocks will be overall neutral. Specifically, feedback will be fluctuating around a mean of 50%. The third, challenge block will be identical to the intervention arm.

3.2 Randomisation

Participants will be assigned to intervention versus control using a random number generation embedded in the task code. For each participant, the javascript command Math.random() will be used to generate a random number between 0 and 1 from a uniform distribution. Participants will then be assigned to condition based on whether the generated number is greater than or less than 0.5.

3.3 Sample size

Using the ANOVA formulation in 6.3, we estimate the required sample size using the following standard formula:

$$n = \frac{\left(z_{1-\alpha/2} + z_{1-\beta}\right)2(1-\rho^2)}{\delta^2}$$

Where α is the probability of Type I error, β is the probability of Type II error, ρ is the correlation of the instrument between baseline and outcome, and δ is the difference between the two groups (effect) powered for.

Setting $\alpha = 0.05$, $\beta = 0.1$ (90% power), and $\rho = 0.12$ (based on our pilot data), for detecting a 0.3 standard deviation difference, in momentary anxiety at the end of the second block (primary outcome) and adjusting for 20% attrition rate, we need a sample size of:

$$n_{infl} = \frac{n}{0.8} = 578$$

3.4 Blinding

This is a double-blind trial. Assessors are blind to treatment allocation; participants are not. Statisticians will also be blinded to allocation as far as possible until after the primary analysis has been agreed. One of the statisticians will attend the Data Monitoring and Ethics Committee (DMEC) and consequently may become unblinded if the committee requires any results (particularly Adverse Events) to be reported separately by study arm.

3.5 Observation times

Data will be collected at the following time-points during the trial:

- baseline
- end of intervention (at the end of the two main blocks)
- at the end of the negative section of the third, 'challenge' block

Not all measures will be recorded at every time point. The data recorded at baseline and post-intervention time points will constitute the full dataset for the purpose of analysis of the primary outcome.

Data will be considered recorded at a given time point (within minutes of it), provided that these

data are collected from each participant within this window. Any participants for whom data are not collected within this time window will be considered missing at that follow-up time for the purpose of the primary analysis.

In the event of any data being collected in error outside of the permitted time windows, the number and percentage of observations excluded for being outside the relevant time window will be summarised separately by study arm for the primary outcome only.

4 Statistical Principles

4.1 Organisation of data and analyses

The SAP will be finalised and approved prior to commencing analysis. The programs and code to be used for statistical analyses will be prepared prior to unblinding as far as possible. Two statisticians will perform the primary analysis independently.

Prior to database lock, basic checks will be performed by the statistician on the blinded data to ensure completeness and accuracy. Outcome variables (primary and secondary), baseline demographic variables and key date variables will be checked for:

- missing values
- values outside an acceptable range
- other inconsistencies

If missing values or other inconsistencies are found, the corresponding data will be sent to the Trial Management Team for checking and will either be corrected, deemed to be missing or confirmed correct, as appropriate.

4.2 Confidence intervals and p-values

All statistical tests will be two-sided. All estimates will be presented with two-sided 95% confidence intervals. For the primary outcome, statistical significance is set at 5%.

4.3 Adherence to intervention

Adherence to the intervention is defined as completion of more than 50% of the trials of the task.

4.4 Analysis populations

The ‘intention-to-treat’ population will include all randomised participants according to the treatment to which they were randomised to receive. Any participants that have withdrawn from the trial, and withdrawn permission to keep and use their data, will be necessarily excluded.

5 Trial Population

5.1 Recruitment and retention

A CONSORT diagram will be presented to provide a detailed description of participant numbers at each time point during the trial. In addition, a table summarising the number of participants who have been lost to follow up at each stage of the trial and reasons for loss to follow up (if supplied) will be presented.

5.2 Baseline characteristics

The demographic information collected at baseline will be presented in a table summarised separately by study arm. Categorical variables will be reported as counts and percentages. Continuous variables will be summarised as either means and standard deviations (SD) or medians and interquartile ranges, depending on the distribution of the data. No statistical tests will be performed to assess baseline differences between study arms. In addition, all baseline outcomes (screening, primary and secondary, see Table 1) will be presented in a table summarised separately by study arm.

The following characteristics will be summarised:

- Age
- Sex
- Gender
- Baseline SPIN scores
- Baseline LSAS scores
- Baseline PHQ-8 scores
- Baseline GAD-7 scores
- Prior diagnoses
- Medication use

6 Analysis

6.1 Outcomes

6.1.1 Primary outcome

Primary Outcome: The primary outcome will be the five-trial average of momentary anxiety by subjective report (visual analogue scale from 0-100, with 0='very relaxed' to 100='very nervous/uncomfortable') at the end of the two 'main' blocks.

6.1.2 Secondary outcomes

Secondary Outcome 1: The primary outcome will be the five-trial average of momentary mood by subjective report (visual analogue scale from 0 -100 for from 0 'very unhappy' to 100 'very happy') at the end of the two 'main' blocks.

Secondary Outcome 2: Resilience to negative feedback within the task. After a short break, during which participants complete a 1-2 minute filler task, we will present participants with the same block of negative feedback and test their resilience to it. Resilience will be tested by taking the 5-trial average mood and anxiety at the end of the negative section of the third block.

6.2 Timing of outcomes

Table 1 provides an overview of primary and secondary outcomes and the time points at which they will be collected.

Table 1: Data collection measures and time points (from Table 1 in protocol).

Measure	Baseline	End of Main blocks	End of negative section of challenge block
SPIN social anxiety measure	✓		
LSAS-SR social anxiety measure	✓		
PHQ-8 depression measure	✓		
GAD-7 Anxiety measure	✓		
Momentary affect	✓	✓	✓

Adverse events Questionnaire			✓

6.3 Primary outcome analysis

We will use the following ANCOVA analysis model

$$Y_{i,END} = \beta_0 + \beta_1 \text{Treatment}_i + \beta_2 Y_{i,0} + \beta_3 \text{Source}_i + \varepsilon_i$$

where $Y_{i,END}$ and $Y_{i,0}$ are the primary outcome at the end of the intervention and at the start of the intervention respectively, Treatment_i indicates participant membership to either the Surprise or the Random arm, and Source_i indicates the 'source' of the participant (0=UCL, 1=online). The measure of interest is estimated through the coefficient of Treatment ($\hat{\vartheta} = \widehat{\beta_1}$) and the test of interest is:

$$H_0: \vartheta = 0 \quad \text{vs.} \quad H_1: \vartheta \neq 0$$

6.3.1 Model checking

All modelling assumptions will be checked, e.g. (approximate) normality and homoscedasticity of residuals.

6.4 Secondary outcome analyses

The effect of the intervention on secondary outcomes will be assessed using analogous methods to those used for the primary outcome. P-values will not be reported for secondary analyses. These analyses will be considered supportive.

6.5 Sensitivity analyses

We will also analyse the primary outcome (data) using a mixed model with random intercepts, i.e. rather than averaging over the five trials.

6.6 Subgroup analyses

We will investigate whether the treatment effect differs between participants recruited from the UCL community and those recruited online. To achieve this, we will re-fit the primary analysis model with an additional interaction term between the terms Treatment and Source, then formally test this term.

6.7 Missing data

Withdrawals from the study, loss to follow up and other missing outcome data will be summarised separately by randomised group. Potential bias due to missing data will be investigated by comparing the baseline characteristics of participants with and without missing values. Depending on the quantity of missing values, predictors of missingness may be identified. We will then perform a sensitivity analysis that includes these predictors of missingness as covariates in the primary analysis

model.
Multiple imputation may also be performed, if deemed appropriate[1]. The imputation model will include outcome data from all time-points, as well as baseline characteristics and any predictors of missingness. The primary analysis model will then be re-fitted using the imputed data.

6.8 Exploratory analyses

We will explore whether total LSAS score and the public speaking factor of the LSAS moderate the primary and secondary outcomes.

6.9 Adverse events

At the end of the task, participants are asked to give some feedback, including to rate how they felt during the intervention, from 0 (Very Upset) to 100 (Very Happy). If participants score <= 50, they are then asked to report what it is that upset them during the task. This feedback will be monitored by the Data Monitoring and Ethics Committee (DMEC).

In all sessions, the following text is displayed at the end of the experiment:
"Thank you very much for participating. If you are currently experiencing any mental health problems you can contact your healthcare provider or charities such as Mind (https://www.mind.org.uk) or Samaritans (https://www.samaritans.org). However, if you provide us with specific issues we could guide you better according to your specific needs—even if you decide not to take part in this study."

All adverse events will be noted by the trial's research assistants in a specific log (including date, and a brief description of the event), which in turn will be shared with the PI who will evaluate them according to their severity. Should an SAE arise, the investigator will notify the Sponsor and the TSC.

6.10 Reporting

Analyses will be reported with regard to the CONSORT checklist[2] and with any particular requirements of academic journals and the funders to which the results of analyses are submitted.

7 Appendix

7.1 List of Abbreviations

ABBREVIATION	DEFINITION
ANCOVA	Analysis of Covariance
ANOVA	Analysis of Variance
CI	Confidence Interval
CONSORT	Consolidated Standards of Reporting Trials
DMEC	Data Monitoring and Ethics Committee
GAD-7	Generalised Anxiety Disorder Questionnaire (7-item)
PHQ-8	Patient Health Questionnaire (8-item)
PI	Principal Investigator
RCT	Randomised Controlled Trial
SAE	Serious Adverse Event
SAP	Statistical Analysis Plan
SD	Standard Deviation
SPIN	Social Phobia Inventory
STAR	Social Training of Affective Response
TSC	Trial Steering Committee

7.2 Coding of Outcomes

7.3 Social anxiety (SPIN)

The Social Phobia Inventory (SPIN; Connor et al., 2000) measures the severity of social anxiety through 17 items, each rated on a 5-point scale, with the response options 0 (“none”), 1 (“a little bit”), 2 (“somewhat”), 3 (“very much”) and 4 (“extremely”). Total scores range from 0 to 68, with scores above 19 indicating a need for further clinical evaluation.

7.4 Social anxiety (LSAS-SR)

The Lebowitz Social Anxiety Scale – Self-Report (LSAS-SR;Fresco et al., 2001; Leigh, et al., 2022) is calculated adding scores of 0, 1, 2 and 3, to the options “Never”, ”Occasionally”, “Often” and “Usually”, respectively. Total score of the LSAS ranges from 0 to 144. Scores of 0-29 indicate no social anxiety; 30-49 mild social anxiety; 50-64 moderate social anxiety; 65-79 marked social anxiety; 80-94 severe social anxiety and 95-144 very severe social anxiety. A score of ≥30

indicates social anxiety and a score of ≥60 indicates generalized social anxiety.

7.5 Depression severity (PHQ-8)

This is calculated by assigning scores of 0, 1, 2, and 3, to the response categories of "Not at all", "Several days", "More than half the days", and "Nearly every day", respectively. PHQ-8 is the same questionnaire as the PHQ-9 (Kroenke et al., 2001) but omits the 9th question about suicidal thoughts and self-harm. The total score for the eight items included ranges from 0 to 24. Scores of 5, 10, 15, and 20 represent cut- points for mild, moderate, moderately severe and severe depression, respectively.

7.5.1 Anxiety (GAD-7)

This is scored in the same way as PHQ-8. GAD-7 (Spitzer et al., 2006) total score for the seven items ranges from 0 to 21. Scores of 5, 10, and 15 represent cut-points for mild, moderate, and severe anxiety, respectively. When screening for anxiety disorders, a recommended cut-point for further evaluation is a score of 10 or greater.

8 References

8.1 Methodology

- Schulz KF, Altman DG, Moher D. CONSORT 2010 Statement: updated guidelines for reporting parallel group randomised trials. BMJ 2010; 340:c332.

8.2 Coding of outcomes

- Connor KM, Davidson JR, Churchill LE, Sherwood A, Foa E, Weisler RH. Psychometric properties of the Social Phobia Inventory (SPIN). New self-rating scale. Br J Psychiatry. 2000 Apr;176:379-86. doi: 10.1192/bjp.176.4.379. PMID: 10827888.
- Fresco, D. M., Coles, M. E., Heimberg, R. G., Liebowitz, M. R., Hami, S., Stein, M. B., & Goetz, D. (2001). The Liebowitz Social Anxiety Scale: a comparison of the psychometric properties of self-report and clinician-administered formats. Psychological Medicine, 31(6), 1025–1035. https://doi.org/10.1017/s0033291701004056
- Kroenke K, Spitzer RL, Williams JB. The PHQ - 9: validity of a brief depression severity measure. Journal of general internal medicine. 2001 Sep;16(9):606-13.
- Leigh, E., & Clark, D. M. (2022). Establishing the psychometric properties of the self-report liebowitz social anxiety scale for children and adolescents in a general population adolescent sample. European Journal of Psychological Assessment, 38(4), 243–250. https://doi.org/10.1027/1015-5759/a000659
- Spitzer RL, Kroenke K, Williams JB, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. Archives of internal medicine. 2006 May 22;166(10):1092- 7.