

RELAX – REducing Levels of AnXiety - a randomised controlled trial of a web-based early intervention targeting repetitive negative thinking (RNT) in pregnant women: an evaluation of its impact on perinatal anxiety and the mechanism of change

Statistical Analysis Plan

Version 1.0: 29/11/2023

Refers to Protocol: 307468_Clinical Investigation Plan_V1.8_28.09.2023_Clean

Version, Version: 1.8

IRAS Reference Number: 307468

ISRCTN: ISRCTN12754931

Trial Statistician: Rose Tinch-Taylor

Signature:  Date: 15/11/2023

Senior Statistician: Prof. Kimberley Goldsmith

Signature:  Date: 15/11/2023

Chief Investigator: Prof. Colette Hirsch

Signature:  Date: 15/11/23

Trial Steering Committee Chair: Prof. Simon Skene

Signature:  Date: 29/11/2023

Amendments to SAP after signoff of version 1.0

Date	SAP version	Protocol version and date	Reason for change

Contents

A) QUANTITATIVE ANALYSIS PLAN....5
1. Description of trial....7
1.1. Principal research objectives to be addressed....8
1.2. Trial design....9
1.3. Internal pilot....11
1.4. Method of allocation of arms....11
1.4.1. Sequence generation and process of allocation....11
1.4.2. Blinding....12
1.5. Duration of the intervention period....14
1.6. Frequency and duration of follow-up....14
1.7. Assessment windows....14
1.8. Data collection....14
1.8.1. Eligibility screening....14
1.8.2. Measures....15
1.8.3. Sample size estimation....19
2. Data Analysis Plan – Data description....19
2.1. Recruitment and representativeness of recruited participants....19
2.2. Baseline comparability of randomised arms....21
2.3. Withdrawal from the intervention and trial....21
2.4. Loss to follow-up and other missing data....21
2.5. Adverse event reporting....21
2.6. Descriptive statistics for outcome measures....21
2.7. Adherence to allocated intervention and intervention fidelity....21
3. Data analysis plan – Inferential Analysis....22
3.1. Definition of primary estimand....22
3.2. Analysis of primary outcome....23
3.3. Analysis of secondary outcomes....26
3.4. Statistical considerations....26
3.4.1. Time points....26
3.4.2. Stratification and clustering....26
3.4.3. Missing items in scales and sub-scales....26
3.4.4. Missing baseline data....27
3.4.5. Missing outcome data....27
3.4.6. Method for handling multiple comparisons....28

3.4.7. Method for handling non-compliance (per protocol/CACE analyses).....28
3.5. Sensitivity analyses.....28
3.6. Planned subgroup analyses.....28
3.7. Exploratory analyses 29
3.8. Interim analyses 29
4. Software.....29
4.1. Data management.....29
4.2. Statistical software.....30
5. References 31
6. Appendix 1 Schedule of Assessments.....33
7. Appendix 2 Brief description of instrument scoring.....34
8. Appendix 3 Manuscript Deviations from or Omissions relative to the Statistical Analysis Plan 35

A) QUANTITATIVE ANALYSIS PLAN

Note differences in this list as compared to the protocol reflect comments on this analysis plan from the study team.

Investigators:
Professor Debra Bick, OBE
Professor of Clinical Trials in Maternal Health. Warwick Medical School, University of Warwick.
debra.bick@warwick.ac.uk

Dr Gertrude (Trudi) Seneviratne, OBE
Consultant Psychiatrist and Clinical Director of Perinatal Mental Health. South London and Maudsley NHS Foundation Trust.
gertrude.seneviratne@slam.nhs.uk

Professor Michelle Moulds
Professor of Psychology. School of Psychology, Faculty of Science, UNSW SYDNEY.
m.moulds@unsw.edu.au

Mary Newburn
Service User Researcher/ Public & Parent Involvement Lead for Maternity, King's College London.
mary.1.newburn@kcl.ac.uk

Jill Newby
Professor & Research Director of the Clinical Research Unit for Anxiety and Depression. School of Psychology, Faculty of Science, UNSW SYDNEY.
j.newby@unsw.edu.au

Chief Investigator:
Professor Colette Hirsch
King's College London.
colette.hirsch@kcl.ac.uk

Trial Coordinator:
Dr Brittannia Volkmer
King's College London.
brittannia.volkmer@kcl.ac.uk

Researchers:
Yogini Sawjani
King's College London.
yogini.sawjani@kcl.ac.uk

Jo Bennett
Lead Research Midwife.
Guy's and St Thomas' NHS Foundation Trust, St Thomas' Hospital.
Jo.Bennett@gstt.nhs.uk

Senior Trial Statistician:
Prof Kimberley Goldsmith
King's College London.

kimberley.goldsmith@kcl.ac.uk

Previous Trial Statistician:
Rose Tinch-Taylor
King's College London.
rose.tinch-taylor@kcl.ac.uk

Current Trial Statistician:
TBD

1. Description of trial

REducing Levels of AnXiety (RELAX) is a two-arm parallel superiority randomised controlled trial of a web-based intervention delivering Cognitive Bias Modification for Interpretation (CBM-I).

RELAX will study a population of pregnant women over the age of 18 who are between 16- and 28-weeks' gestation. To be eligible for the study, the women will have high levels of repetitive negative thinking (RNT) and low to moderate anxiety. The study will be conducted within 2 NHS sites, King's College Hospital (KCH) and Guy's and St Thomas' NHS Foundation Trust (GSTT), and 1 non-NHS site (King's College London (KCL)), with the non-NHS site recruiting participants from social media, pregnancy community organisations, charities, and support groups.

The study will test the efficacy of a web-based psychological intervention for perinatal anxiety, RELAX. The online platform uses 12 CBM-I training sessions to train pregnant women to interpret uncertain/ambiguous information in realistic and more positive ways. In each session, users listen to 30 short, ambiguous scenarios that can be interpreted in both negative and positive ways. Some of the scenarios are resolved in a positive way by the platform and the user is required to resolve some of them in a positive manner themselves. Users then imagine themselves in the positive outcome for each scenario. RELAX aims to reduce anxiety (the primary outcome) during the perinatal period. The RELAX platform has been developed based on a series of research experiments that have shown a similar web-based CBM-I intervention (1) to have positive effects in non-pregnant people with clinical anxiety and/or depression. All study participants will receive usual care (UC), which typically involves monitoring by maternity services and contact with a health visitor. Women may also be offered information on self-referral to local psychology services where they will be put on a wait list or be offered generic interventions (e.g., group treatment, computerised CBT). Women in the intervention arm will receive UC plus the RELAX intervention (referred to in this document as RELAX).

Women will complete assessments online at: (1) T0 baseline assessment pre-randomisation (16-28 weeks gestation), (2) T1 assessment 4-weeks post-randomisation (20-32 weeks gestation), (3) T2 8-weeks post-randomisation (24-36 weeks gestation), and (4) T3 36-weeks post-randomisation (12-24 weeks post birth). For participants randomised at 16-28 weeks gestation, and pregnancy assumed a predictable 40-week duration, 36 weeks after randomisation will be 12-24 weeks after birth.

To test the efficacy of RELAX, we will assess the level of anxiety for women at T2 (8-weeks after randomisation), when women are between 24-36 weeks gestation. This will be

achieved by estimating the mean difference in anxiety scores using the 7-item Generalized Anxiety Disorder rating scale (GAD-7) between the RELAX and UC arms (primary outcome) whilst adjusting for baseline GAD-7.

Further, we will determine whether the difference in interpretation bias between the RELAX and UC arms measured using the Recognition Test (RT) at T1 (4-weeks post randomisation), when women are between 20-32 weeks gestation mediates the relationship between the RELAX intervention and:

i. Anxiety **before birth** at 8-weeks post randomisation when women are between 24-36 weeks gestation.
ii. Anxiety **after birth** at 36-weeks post randomisation when women are between 12-24 weeks post-birth.

Secondary outcomes will assess the effect of RELAX versus UC on 1) anxiety **after birth** (36-weeks post-randomisation); 2) repetitive negative thinking, general depression, worry, perinatal specific anxiety and depression, and work and social functioning **before** (8-weeks post-randomisation) **and after** (36-weeks post-randomisation) birth.

We will also assess the effect of the intervention in the population of compliers via a complier average causal effect estimate (CACE) analysis.

This statistical analysis plan (SAP) solely describes the quantitative analyses that will be carried out by the trial statistician.

1.1. Principal research objectives to be addressed

This objectives section is taken directly from the trial protocol with some minor modifications.

The overall objective of the trial is to see if a new online training (RELAX) plus usual care leads to lower anxiety in comparison to usual care only in pregnant women and new mothers with high levels of RNT and a low to moderate level of anxiety.

Primary objective:

1. The primary clinical objective is to assess whether at-risk pregnant women who receive RELAX alongside receiving usual care experience less anxiety at 8-weeks post randomisation **before birth**, than at-risk pregnant women who only receive usual care.
2. The primary mechanistic objective is to determine whether interpretation bias is the mechanism of action on changes in anxiety of the RELAX intervention plus usual care, in comparison to at-risk pregnant women who only receive usual care.

Secondary objectives:
As this is an efficacy trial, the main secondary analysis will be a complier average causal (CACE) estimate to compare outcomes for women in the RELAX arm who complied, with individuals in the control arm who would have complied if they had been randomised to the treatment arm.

A complier is defined as a participant who has completed all 30 trials within a session, for at least 10 out of the 12 RELAX sessions. A trial is defined as a scenario, followed by a comprehension question and either a positivity or vividness rating.

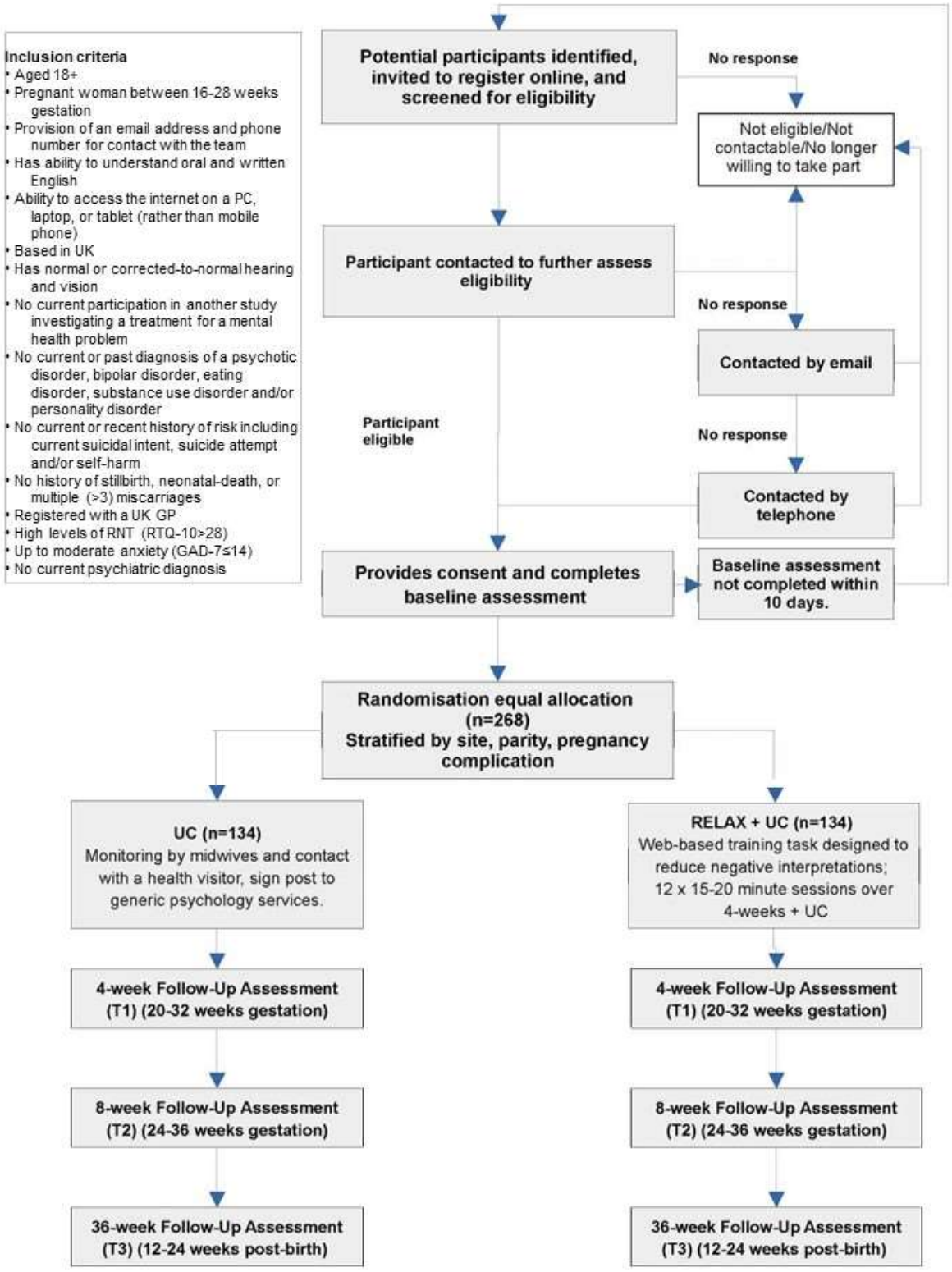
Further, to investigate whether, in comparison to pregnant women who only receive UC, those who complete RELAX alongside receiving UC have i) less anxiety **after** birth ii) lower levels of repetitive negative thinking, less general depression, lower levels of trait worry, less perinatal specific anxiety and depression, and better work and social functioning **before and after** birth.

More details of the intervention and study design are provided in the protocol, 307468_Clinical Investigation Plan_V1.8_28.09.2023_Clean Version [version 1.8].

1.2. Trial design

The RELAX trial is a two-arm superiority parallel individually randomised controlled trial. The duration of the intervention is four weeks, and measures are scheduled at baseline (within 10 days of completing the screening questionnaire), and at 4-weeks, 8-weeks, and 36-weeks post randomisation. Figure 1 provides an overview of the study schedule.

Figure 1: Trial design flow diagram



1.3. Internal pilot

An internal pilot will assess recruitment and T2 and T3 assessment completion rates to determine the feasibility of trial completion. The pilot is purely a term used to describe milestone review points of recruitment and assessment completion rates to establish ongoing viability. No additional data will be collected for this purpose, no separate variables will be compiled.

The internal pilot assessment for progression criteria will be 16 months after the start of the trial (5 months' worth of recruitment, 10 months to complete all assessments, 1 month leeway for late assessments).

Progression to the full trial will be justified if recruitment is on average at least 12 per month and assessment completion rates for T2 and T3 are at least 65% (sample size accounts for 35% attrition). 95% CI for recruitment and completion rates will be presented. If progression is not justified, we will consult the TSC for their recommendations and consult with NIHR and co-sponsors regarding further steps potentially including stopping all further recruitment and assessments.

1.4. Method of allocation of arms

1.4.1. Sequence generation and process of allocation

Avegen, a digital healthcare company who deploy digital health applications through cloud-hosted, product development platforms (**www.avegenhealth.com**), will develop the online randomisation system in collaboration with the statisticians and the CI.

Participants will be randomised in a 1:1 ratio. Avegen will construct a randomisation list for each combination of strata using random permuted blocks with varying block sizes. The investigators, including the Chief Investigator and statisticians, will not have access to the allocation sequence. The Avegen platform will gather information on the three stratification variables: recruitment site (KCH, GSTT, KCL), parity (0 vs not 0 previous births at 24+ weeks), and pregnancy complications (no pregnancy complications vs current and/or past pregnancy complications) and use the next allocation from the appropriate list to assign participants. Please see Section 3.6 for a breakdown of pregnancy complications. Two extra recruitment sites will be allowed for in the randomisation list that we will designate as "site 4" and "site 5" in case we find we need to seek to randomise from outside the three main sites. Stratified randomisation will take place after an eligible participant provides informed consent and completes the baseline assessment. The detailed randomisation process is as follows:

i. The participant submits the screening and eligibility questionnaire online, which also gathers the information needed to classify participants based on the stratification variables.
ii. The researcher conducts the study outline call, confirms the stratification variables, and confirms that the participant can be randomised by marking them on the platform as 'eligible to be included in the study.'

iii. The researcher sends the link to the participant to set a password and complete the baseline assessment. The platform will then generate a unique identifier that will also be used to allow linkage of the questionnaire responses.
iv. Following submission of the full, complete baseline assessment the participant will be automatically assigned to a trial arm depending on their strata and the block sequence.
v. The platform will display the assigned arm to the participant. Researchers will remain blinded to the complete randomisation list.

Participant names, phone numbers, and email addresses will be stored within the Avegen platform but will not be used when randomising participants. To ensure data protection, trial statisticians will not be able to view this personal data.

A full audit trail of data entry regarding randomisation will be automatically date and time stamped, alongside information about the researcher making the entry within the system.

1.4.2. Blinding

Please see Table 1. Due to the nature of the study, participants will be unblinded to their allocation arm throughout their full participation in the trial. All participants will receive the same number of automatic emails and phone call prompts to complete each assessment, regardless of their allocation arm.

In the event of clinical need, the Chief Investigator will be unblinded at the individual level (i.e. will have access to the allocation of an individual participant) but will have no access to the randomisation list or the data summarised at arm-level.

The senior trial statistician will not have access to the complete randomisation list or direct access to the outcome data in the Avegen platform at any point in the trial to remain fully blinded. Full unblinding will occur when reviewing the first draft of the statistical report. The trial statistician will be blinded until the initial version of the statistical analysis plan is signed-off, after which they will become unblinded. The rationale for being unblinded is to facilitate their monitoring of the platform usage/intervention-related data. Alongside the data monitoring committee (DMEC) members, who will be partially blinded, the trial statistician will be the only person with access to summarised data by study arm before circulating the statistical report. The study team will take measures to restrict access to adherence data to avoid accidental unblinding of the trial statistician prior to the first statistical analysis plan sign off.

The trial coordinator and researchers will remain unblinded throughout the trial.

Table 1: Blinding status of personnel

Roles	Individual Level	Group Level	Method of blinding OR justification for unblinding
Trial Coordinator	U	B	No access to randomisation list or data summarised at group-level
Research Assistant	U	B	No access to randomisation list or data summarised at group-level
Research Midwife	U	B	No access to randomisation list or data summarised at group-level
Other members of study team	U	B	No access to randomisation list or data summarised at group-level
Study participants	U	B	No access to randomisation list or data summarised at group-level
Trial Statistician	B	U	Access to full randomisation list required for data monitoring & analysis. NB only unblinded after an initial draft of the SAP has been signed.
Senior Statistician	B	B	No access to full randomisation list *
Chief / Co / Principal Investigators	U	B	No access to randomisation list or data summarised at group-level
(Independent members of) DMEC	B	U/P	The level of DMEC blinding will be at their discretion but will likely see data split by group at least at a partially blinded level
Trial Steering Committee (TSC)	B	B	No access to randomisation list or data summarised at group-level

U = unblinded, B = blinded, P = partially blinded (i.e., see data split by groups labelled as A/B).
*After data has all been collected, the database locked, the data have been analysed and a first draft of the statistical report has been prepared, the senior statistician will become unblinded to carry out final checks on the analysis code and statistical report.

1.5. Duration of the intervention period

Participants in the intervention arm will complete RELAX, which consists of 12 web-based CBM-I training sessions (of 15-20 minutes duration), over a period of 4 weeks alongside receiving usual care. One session is composed of 30 trials. A trial includes listening to a scenario, answering a comprehension question, and answering a vividness or positivity question.

1.6. Frequency and duration of follow-up

The aim will be for randomisation to occur instantly on completion of the baseline measures. Follow-up assessments will take place 4-weeks + 1day, 8-weeks + 1day, and 36-weeks +1day post-randomisation. Please refer to the Appendix 1 Schedule of Assessments for a more comprehensive description of the assessments planned at each assessment time point.

Each participant will remain in the trial for approximately 36 weeks. Participants in the intervention arm will access the intervention on the RELAX platform for four weeks from randomisation. Given that this is a fully web-based trail recruiting across the UK, research data will also be gathered remotely (via an online platform, potentially with some information gathered via phone calls).

1.7. Assessment windows

The following assessment windows apply to the assessments carried out in the trial:

- Pre-randomisation (baseline) assessment: assessment window 0 to 10 days from completing screening questionnaire.
- 4-week + 1day (29 days) post-randomisation assessment: assessment window +21-days, i.e. 50 days post-randomisation.
- 8-week + 1day (57 days) post-randomisation assessment: assessment window +21-days, i.e. 78 days post-randomisation.
- 36-week + 1day (253 days) post-randomisation assessment: assessment window +56-days, i.e. 309 days post-randomisation.

1.8. Data collection

1.8.1. Eligibility screening

Inclusion criteria

1. Pregnant women between 16-28 weeks gestation
2. Aged 18 years or older
3. Based in the United Kingdom
4. High levels of RNT (Repetitive Thinking Questionnaire (RTQ-10) (trait) ≥ 28)
5. Anxiety up to only a moderate level (GAD-7<15)

6. Ability to understand oral and written English
7. Normal or corrected-to-normal hearing and vision
8. Ability to access the internet on a PC, laptop, or tablet (rather than mobile phone)
9. Provision of an email address and phone number for contact with the team.

Exclusion criteria:

1. Current psychiatric diagnosis (this will be assessed via the clinical interview schedule revised (CIS-R; Lewis & Pelosi, 1992) which participants will complete as part of the screening questionnaire as well as via general questions about a current diagnosis of a psychotic disorder, eating disorder, substance use disorder or personality disorder)
2. Past diagnosis of a psychotic disorder (e.g., schizophrenia), bipolar disorder, an eating disorder, a substance use disorder (e.g., alcohol dependence) and/or a personality disorder (e.g., borderline personality disorder)
3. Current or recent history of risk: current suicidal intent (PHQ-9 item 9>1), suicide attempt <2 years and/or self-harm <1 year
4. History of stillbirth, neonatal-death, or multiple (i.e., ≥3) miscarriages
5. Current participation in another study that is investigating an intervention for a mental health problem
6. Not registered with a GP in the UK

1.8.2. Measures

This section contains an overview of those measures relevant to the quantitative analysis only. Please also refer to the schedule of assessments (Appendix 1 Schedule of Assessments) and a brief description of instrument scoring (Appendix 2 Brief description of instrument scoring) for further description of timings of assessments and information on scoring algorithms.

1.8.2.1. General description

Anxiety. Assessed using the Generalized Anxiety Disorder (GAD-7) measure. A well-validated and widely used 7-item measure of anxiety (1). The GAD-7 is the primary clinical outcome measure.

Interpretation bias. Assessed using the Recognition Test (RT) (2). This measure, which assesses interpretation bias, has been used widely in the anxiety and depression literature as well as this team's previous work with pregnant women (3).

Depression. Assessed using 9-item Patient Health Questionnaire (PHQ-9). This is the 9-item depression module from the full PHQ. There is strong evidence for the validity of the PHQ-9 as a brief measure of depression severity. A score of 10 to 27 indicates moderate to severe depression (4).

Repetitive Negative Thinking (RNT). Assessed using the RTQ-10 (trait) (5), a validated transdiagnostic measure of RNT.

Worry. Assessed using the Penn State Worry Questionnaire (PSWQ, (6)), an established measure of trait worry (i.e., general tendency to worry). This measure is included because worry is the primary form of RNT in relation to anxiety.

Perinatal depression. Assessed using the Edinburgh Postnatal Depression Scale (EPDS, (7)); the most common measure for perinatal depression. Depression is known to be highly associated with anxiety, so this measure has been included as it is relevant to the target population.

Perinatal specific anxiety. The Perinatal Anxiety Screening Scale (PASS, (8)) measures perinatal specific anxiety and was selected because it is the most appropriate perinatal anxiety scale available. This was chosen as a secondary measure because neither the PASS nor any other perinatal anxiety scales are used in IAPT services (Improving Access to Psychological Therapy) or the NHS (National Health Service). Furthermore, the PASS has less validation than the GAD-7, and is not yet published as a clinically meaningful change metric.

Work and social functioning. Assessed using the Work and Social Adjustment Scale (WSAS) to gauge the impact of a person's mental health difficulties on their ability to function in terms of work, home management, social leisure, private leisure and personal or family relationships. It is assessed via a 5-item scale that is commonly used in the NHS and IAPT. The WSAS exhibits good internal consistency and test-retest reliability as a measure of disorder-related functional impairment in adults with anxiety disorders (9). If participants are not working, the item related to work will not be applicable. In this case, the WSAS score will be prorated: the mean of available scores will be imputed for the NA item, and the total score calculated.

1.8.2.2. Baseline measures

Note some measures are gathered at the screening visit (see Appendix 1).

- Demographics/Participant characteristics:
    - Age
    - Sex at birth
    - Ethnicity
    - Religion
    - Relationship status
    - Highest level of education
    - Employment status
    - Pre-existing physical health condition
    - If any exist, type of pre-existing physical health condition
    - Number of children (not including current pregnancy)
    - Weeks of gestation
    - Social support
    - Past/current diagnosis of mental health condition
    - Currently receiving psychological treatment
    - Currently taking medication for psychological disorders
    - History of consulting a mental health professional about mental health problems (will be coded as yes if a) answer yes to spoken to a mental health professional question, b) answer yes to one or more of the questions about

whether have been diagnosed with a mental health condition, received medication for a mental health condition or received psychological treatment even if the answer to the spoken to a mental health professional question is no).

- Stratification factors (site, parity, pregnancy complications)
- Anxiety (GAD-7 scale)
- Interpretation bias (RT)
- Depression (PHQ-9 scale)
- Repetitive Negative Thinking (RTQ-10 trait scale)
- Worry (PSWQ scale)
- Perinatal depression (EPDS scale)
- Perinatal anxiety (PASS scale) global score
- Work and Social Functioning (WSAS scale)

1.8.2.3. Primary outcome measure

- GAD-7 at 8-weeks post-randomisation

1.8.2.4. Primary mechanistic measures

- Mediator: Interpretation Bias measured using the Recognition Test (RT) at 4-weeks post-randomisation
- Outcomes: GAD-7 at 8-weeks

1.8.2.5. Secondary outcome measures

- GAD-7 at 36-weeks post-randomisation
- PHQ-9 at 8-weeks post-randomisation
- PHQ-9 at 36-weeks post-randomisation
- RTQ-10 (trait) at 8-weeks post-randomisation
- RTQ-10 (trait) at 36-weeks post-randomisation
- PSWQ at 8-weeks post-randomisation
- PSWQ at 36-weeks post-randomisation
- EPDS at 8-weeks post-randomisation
- EPDS at 36-weeks post-randomisation
- PASS global score at 8-weeks post-randomisation
- PASS global score at 36-weeks post-randomisation
- WSAS at 8-weeks post-randomisation
- WSAS at 36-weeks post-randomisation

1.8.2.6. Secondary mechanistic measures

- (Mediator: Interpretation Bias measured using the Recognition Test (RT) at 4-weeks post-randomisation)
- Outcomes: GAD-7 at 36-weeks post-randomisation

1.8.2.7. Other measures

- Pregnancy complication as yes/no in the last month at 4- and 8-weeks post-randomisation
- Birth complication as yes/no at 36-weeks post-randomisation

- Currently receiving psychological treatment as yes/no at 4-, 8- and 36-weeks post-randomisation
- Currently taking medication for psychological disorders as yes/no at 4-, 8- and 36-weeks post-randomisation

We will also summarise the following variables as per Section 2.6 Descriptive statistics for outcome measures.

- All of the secondary measures listed just above at 4-weeks post-randomisation
- Pregnancy complications measured at 4- and 8-weeks post-randomisation
- Birth complications measured at 36-weeks post-randomisation (may be collected at 4- or 8-weeks if indicate on pregnancy complications form that they have had their baby by those time points)
- Currently receiving psychological treatment at 4-, 8- and 36-weeks post-randomisation
- Currently taking medication for psychological disorders at 4-, 8- and 36-weeks post-randomisation
- GAD-7 anxiety clinical reliable change from baseline at 4-, 8- and 36-weeks post-randomisation categorised as follows (10, 11):
    - Clinical improvement – a decrease of four points or more
    - No change – increase or decrease of 0 – 3 points
    - Clinical deterioration – an increase of four points or more

For those in the RELAX intervention arm (see Section 1 for a brief description and the protocol for more detail):

- Adherence data including:
    1. Number of sessions completed
    2. Number of trials completed per session
- Session data
    1. Comprehension question accuracy for each trial within the session
    2. Positivity and vividness scores (with each trial including either a positivity or a vividness rating) for each trial within the session
- Intervention expectancy (collected at the end of session 1)
    1. How logical the intervention will be.
    2. How useful the intervention will be.
- Intervention acceptability (collected at 4-weeks post-randomisation)
    1. How logical the intervention was.
    2. How useful the intervention was.
    3. How much confidence the participant has in recommending it to a friend

Please see section 1.5 Duration of the intervention period for session and trial definitions.

1.8.2.8. Adverse Events

Potential adverse events will be reported by participants at each assessment time-point (4-weeks, 8-weeks, and 36-weeks post-randomisation). Researchers will then assess whether these can be classified as Adverse Events and if so, will record these in the separate researcher-entered adverse events form in the main database.

1.8.3. Sample size estimation

A total of 268 women will be recruited into the study.

For a two-sided independent samples *t*-test, alpha = 0.05, assuming one baseline and three post-randomisation measures with correlation rho = 0.5 (conservative based on baseline – T2 estimate from Hirsch, Krahe et al., (12) of rho = 0.42), deflation factor of 0.5 (Machin et al., (13)), and accounting for 35% attrition (based on online CBT: Van Ballegooijen et al., (14)), we will need 268 participants in total (134 per arm) to have 90% power to detect a GAD-7 effect size of 0.35. This is powered slightly conservatively, since natural anxiety trajectories in the perinatal period may be more variable, leading to greater standard deviations compared to non-pregnant populations included in our prior research – in which the T2 effect size for GAD-7 was 0.40 (adjusted mean difference of 2/pooled SD of 5 at T2, Hirsch, Krahe et al., (12)). The previous study utilised a task matched control which also led to reductions in anxiety. The current study with a usual care comparator (appropriate for an efficacy trial) may well yield greater differences. This may be particularly pertinent for the current population, for whom RNT is likely to lead to escalating anxiety for those individuals not offered the active intervention during this potentially stressful time in their lives. We consider a four point between-condition difference to be clinically meaningful on the basis that four points within-person change on the GAD-7 has been defined as clinically meaningful (Toussaint et al., 2020).

Mechanistic analysis sample size

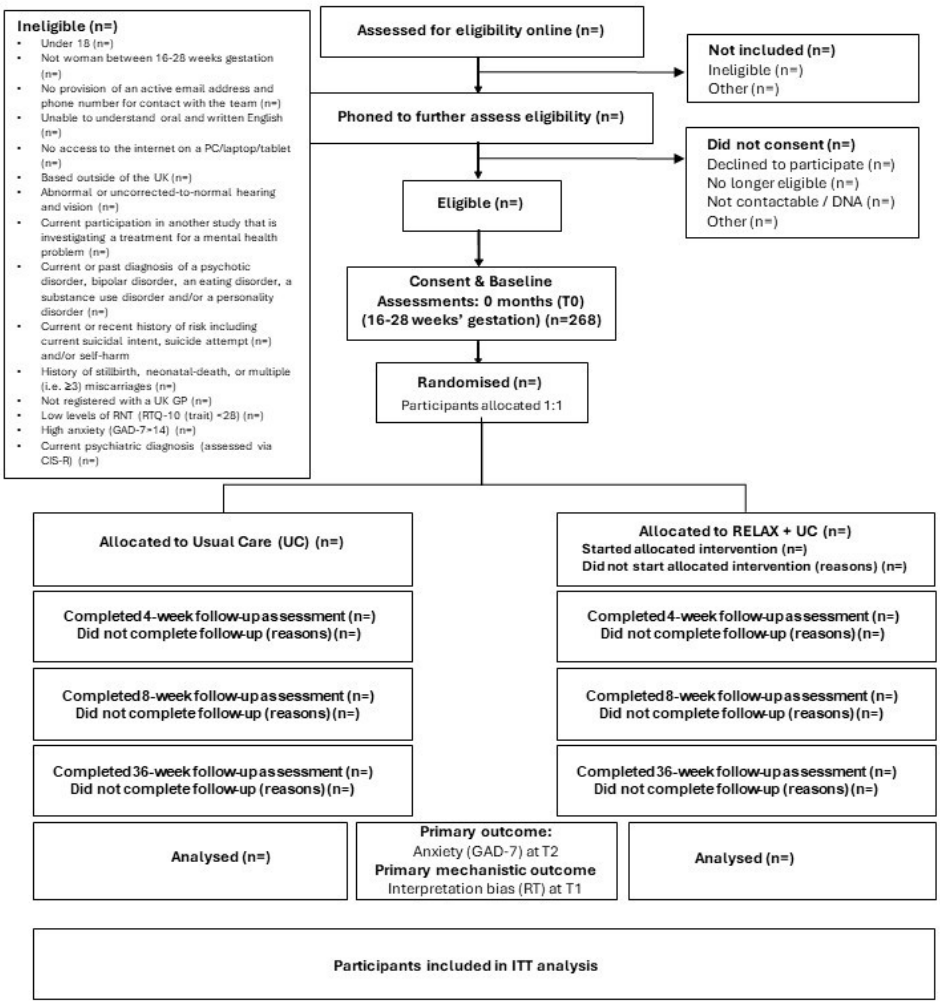
Using estimates of the mediation parameters from Hirsch, Krahe et al. (2021, (12)) to assume standardised estimates of the *a* (effect of intervention on mediator), *b* (effect of mediator on outcome) and *c'* (direct effect of intervention on outcome) paths in a mediation model of approximately 0.5, 0.2 and 0.15, and the methods of Thoemmes et al. (2010,) to calculate appropriate parameters for Monte Carlo simulations with 10K repetitions to calculate power in Mplus (Thoemmes et al., 2010), a total sample size of 268 would give > 99% power to detect the mechanistic action of the intervention on interpretation bias (*a* path), and 83% power to detect the mediated effect of the intervention on GAD-7 via interpretation bias.

2. Data Analysis Plan – Data description

2.1. Recruitment and representativeness of recruited participants

We will construct a CONSORT flow chart (15) to describe the number of: eligible participants, participants invited to the trial, participants excluded with reasons, participants that consent, participants randomised, participants in each arm. Finally, we will explicitly describe the number of participants included in the analyses. Please see Figure 2 for an example trial CONSORT diagram.

Figure 2: Template CONSORT diagram for RELAX trial



2.2. Baseline comparability of randomised arms

Summary statistics describing participants overall and by intervention arm will be reported for the list of variables in Section 1.8.2.2. We will not perform formal inference tests or present p-values for the differences between the arms on baseline variables. We will describe categorical variables using frequencies and proportions. In the case of numerical variables, we will provide means and standard deviations (SD), and medians and interquartile ranges (IQR). Informal comparisons between the means and medians will help assess whether variables are normally distributed (i.e., if within 10% of one another, variable can be assumed to be normally distributed). We will also assess normality visually using quantile-quantile plots and histograms.

2.3. Withdrawal from the intervention and trial

We will separately summarise the number and proportion of participants who actively withdraw from the trial and their reasons for withdrawal, overall and by arm.

2.4. Loss to follow-up and other missing data

The number and proportion of participants missing each primary and secondary outcome variable will be presented by intervention arm and overall, at each time point.

2.5. Adverse event reporting

Adverse events, serious adverse events, adverse device effects (i.e., adverse events deemed to be related to the RELAX intervention), serious adverse device effects, and device deficiencies will be summarised overall and by intervention arm as counts of events/effects and counts of people who have had events/effects. Adverse events will be described by body system code and outcome. Adverse events will only be presented descriptively. We will aim to separate events and device effects, but note that these may not be mutually exclusive. We will indicate if this leads to double counting across events and device effects.

2.6. Descriptive statistics for outcome measures

We will summarise primary and secondary outcomes, including mechanistic outcomes, overall and by trial arm at 4-weeks, 8-weeks, and 36-weeks post-randomisation, where available. We will report categorical variables as frequencies and proportions, numerical variables using means and SD, and medians and IQR. We have defined outcome measures above (please refer to Sections 1.8.2.3, 1.8.2.4, 1.8.2.5, 1.8.2.6).

2.7. Adherence to allocated intervention and intervention fidelity

For a participant in the RELAX arm to be considered as adhering to the intervention, they must complete all of the 30 trials in a session for at least 10 out of the 12 RELAX sessions. We will report the number and proportion of participants meeting the adherence criterion out of all of those allocated to the intervention arm, and out of those who started at least one session.

We will describe intervention expectancy and acceptability variables in the intervention arm.

We will also summarise RELAX usage variables (please see Section **Error! Reference source not found.**) to provide information about usage pattern. We will describe these variables for participants in the RELAX arm using means and SD, medians and IQR, or frequencies and proportions and as appropriate.
The following usage variables will be described:

- Number and % of started sessions out of 12
- Number and % of completed sessions out of 12: a completed session is defined as completing all 30 trials within the session.
- Receiving: no intervention (0 sessions), partial adherers (1-9 sessions), full adherers (≥ 10 sessions)
- Number and % of trials completed per session by session number: statistically a completed trial will be where a comprehension question is answered, and a vividness or positivity question is answered.
- Mean (SD) and median (IQR) of mean accuracy rates of comprehension questions across trials by session number.
- Mean (SD) and median (IQR) of mean positivity and vividness scores across trials by session number.
- Median (IQR) ratings of how logical and useful women expected the intervention to be, collected at the end of session 1.
- Median (IQR) ratings of how logical and useful women found the intervention to be at the end of the 4-week intervention.
- Median (IQR) ratings of how confident women were to recommend RELAX to a friend at the end of the 4-week intervention.

3. Data analysis plan – Inferential Analysis

We will perform all analyses following the intention to treat (ITT) principle as much as possible (16). This means all randomised participants (eligibility as defined in section 1.8.1) are included in the analysis and analysed in the arms to which they were randomised, regardless of which intervention they received post-randomisation.

As the primary and secondary outcomes are continuous, we will report the mean differences in instrument scores between the intervention (RELAX) and control (UC) arms and their associated two-sided 95% confidence intervals, adjusting for the baseline of the outcome. These results will be estimated using linear mixed-effects models with all post-randomisation measures as dependent variables. Sections 3.2 and 3.3 describe the models we will use for the primary and secondary outcomes, respectively. The significance level will be 5% (two-sided) for the primary outcome as well as for the secondary outcomes.

3.1. Definition of primary estimand

The primary estimand is the baseline-adjusted mean difference (and associated confidence interval) in participant-reported GAD-7 (also adjusted for stratification variables site, parity, pregnancy complications, and possibly baseline variables that predict missing data); measured at 8 weeks post-randomisation; between the web-based psychological

intervention for perinatal anxiety - RELAX arm and a usual care only arm; in pregnant women with high levels of repetitive negative thinking (RNT) and low to moderate anxiety; with all measures included regardless of the occurrence of any intercurrent events (i.e., treatment policy/intention to treat approach).

The estimand is described by the following attributes:

- Population: Pregnant women with high levels of repetitive negative thinking (RNT), low to moderate anxiety, and who meet the trial inclusion/exclusion criteria.
- Endpoint (primary): Anxiety (measured using the GAD-7) at 8-weeks post-randomisation.
- Treatment condition: REducing Levels of AnXiety (RELAX), a web-based intervention delivering Cognitive Bias Modification for Interpretation (CBM-I), compared to usual care (UC) regardless of discontinuation for any reason (treatment policy strategy).
- Population-level summary: Adjusted Mean Difference on the raw scale, adjusted for baseline GAD-7 and the site, parity and pregnancy complications stratification variables, with 95% confidence intervals and p-value. If there are baseline variables that predict missing data, these will also be adjusted for.
- Handling of intercurrent events: treatment policy, i.e., the intervention effect will be estimated regardless of any intercurrent events/the measured values of the variables will be used regardless of any intercurrent events.

3.2. Analysis of primary outcome

Clinical Efficacy

The primary outcome for RELAX is the mean difference between the GAD-7 measure between arms at 8-weeks post-randomisation. This will be estimated using a mixed-effects linear analysis of covariance (ANCOVA) model with repeated measures and a random intercept at the participant level. We will include as dependent variables the GAD-7 scores at 4-weeks, 8-weeks (primary outcome) and 36-weeks post-randomisation. The independent variables included will be trial arm, time, baseline anxiety (GAD-7), weeks of gestation at baseline (a pre-specified baseline variable), the three randomisation stratification variables, and an interaction term for trial arm and time. Time will initially be entered as a categorical variable (i.e. time point) and we will explore whether random coefficients are warranted by entering categorical random effects time terms and using a likelihood ratio test comparing to the random intercept only model. If the likelihood ratio test for the comparison is statistically significant, we will retain the random coefficients. If we experience model fit issues due to the use of categorical time, we will explore time since randomisation as a continuous variable and assess the need for a random slope over time in a similar manner. Furthermore, if any baseline variables are found to predict missing outcome data (Section 3.4.5), we will include them as covariates in our model. The

interaction term between trial arm and time will be used to obtain the RELAX vs UC mean difference at 8-weeks post-randomisation.

Mechanistic analysis

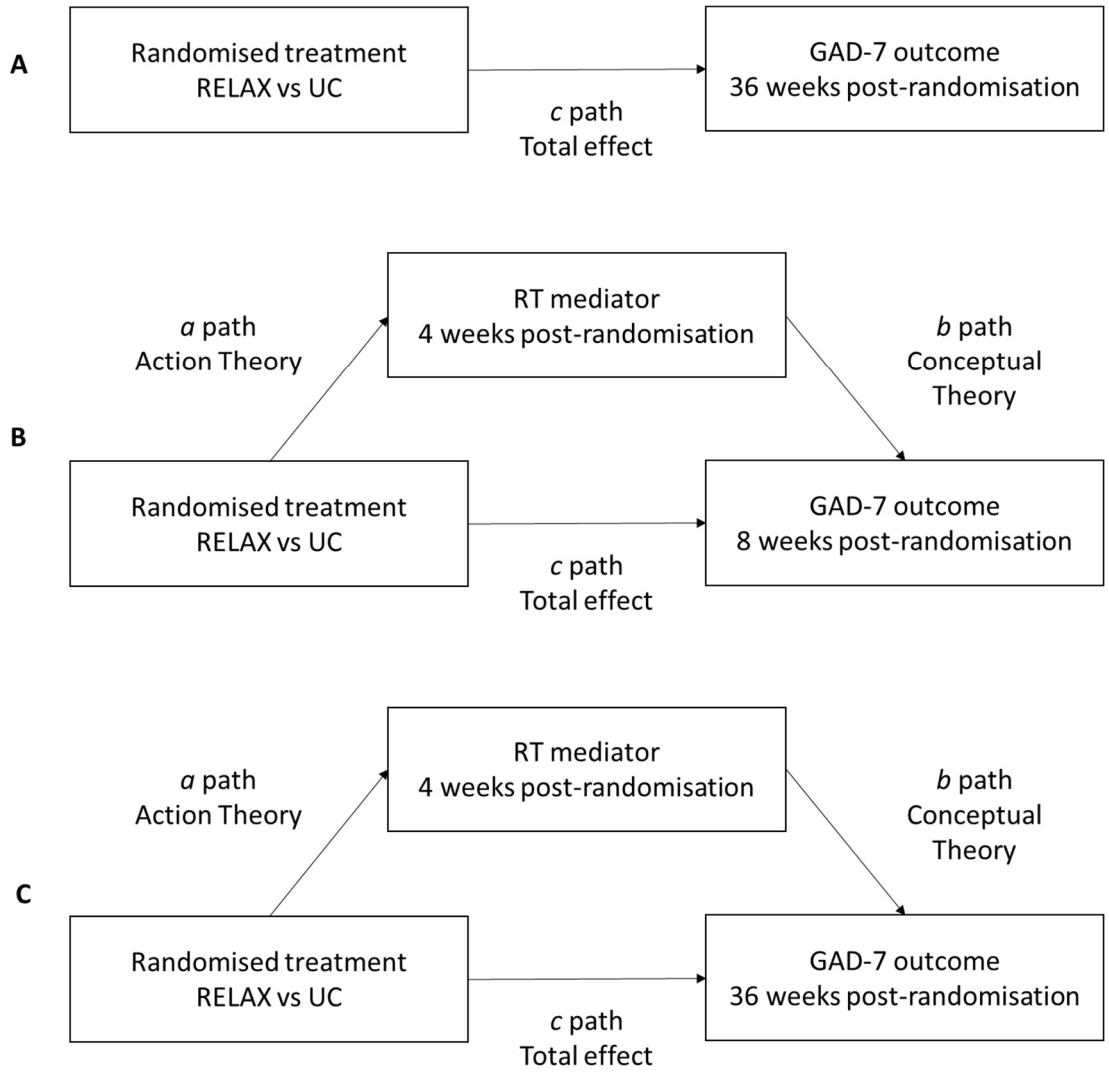
To estimate to what extent the intention to treat RELAX vs UC intervention effect on anxiety (GAD-7) at 8-weeks (primary) or 36-weeks is transmitted via interpretation bias between arms (measured using the Recognition Test) at 4-weeks, we will use structural equation modelling with full information maximum likelihood estimation, e.g., Stata sem command using the method (mlmv) option, or something similar, to fit mediation models. We will fit two separate structural equation models for the 8-week and 36-week GAD-7 outcome measures, each fitting two equations: one with the 4-week interpretation bias (RT) mediator as the dependent variable and the other with the GAD-7 outcome as the dependent variable with the RT mediator as an independent variable (Figure 3B and C). Both equations for both models will have arm as a binary independent variable and include the three randomisation stratification factors and the weeks of gestation at baseline variable as independent variables. Baseline GAD-7 and interpretation bias (RT) measures will also be included as independent variables in both the mediator and outcome models (not shown in Figure 3 for simplicity). We will likely also include as independent variables any other baseline variables found to predict missing outcome data as per Section 3.4.5.

We will initially determine the total effects of intervention on the outcome for each of these models (an example for GAD-7 at 36 weeks is shown in Figure 3A), because these individual time point effects might differ somewhat from the effects from the main mixed effects analysis models incorporating both GAD-7 outcome time points. We will obtain these total effects using linear regression with GAD-7 at the appropriate time point and with the independent variables as described just above, barring the interpretation bias mediator variable. We will then fit the mediation models shown in Figure 3B and C, and report the following from each, all along with associated 95% CI: *a* paths; the effect of intervention arm on the interpretation bias mediator as mean differences between the arms, the *b* paths as the change in the GAD-7 outcome for a one point change in the interpretation bias mediator; and the c' paths; as mean differences in GAD-7 between the arms taking the mediator into account. We may consider using standardised variables for the mediation modelling; this would be all variables standardised to baseline (i.e. take each variable, create a new variable by subtracting its baseline mean and dividing by its baseline SD).

The indirect (or mediated) effect will be calculated in each case by multiplying the *a* and *b* paths together. For the indirect effect, we will present the percentile bootstrap CI from bootstrapping with 1000 repetitions (17). The analysis will therefore allow us to partition the total effect of the RELAX intervention on GAD into the portion that is transmitted via interpretation bias, with the remaining portion being via the direct effect. We will also

present this proportion of the total effect that mediated for each model, calculated as the indirect effect estimate divided by the total effect estimate, multiplied by 100.

Figure 3: Mechanistic analysis models



3.3. Analysis of secondary outcomes

Clinical efficacy

We will test the intervention effects on GAD-7 scores at 36-weeks post-randomisation, as well as on the scores of the other clinical instruments at 8- and 36-weeks post-randomisation (please see Section Secondary outcome measures). The mean difference in GAD-7 scores between arms at 36-weeks post-randomisation will be estimated using the clinical efficacy model described in Section 3.2. The mean differences between arms for the scores of the other clinical instrument secondary outcomes at 8- and 36-weeks post-randomisation will be estimated using similar mixed-effects linear ANCOVA models as the one described in Section 3.2. We will use all the available post-randomisation scores for each instrument as dependent variables (4-, 8-, and 36 weeks post-randomisation) and include the same covariates as the ones described for the primary outcome in section 3.2 with the relevant baseline measure substituted in each case, as well as including baseline measures predictive of missing outcome data (see Section 3.4.5).

CACE analysis

As a main secondary analysis, we have also specified a Complier Average Causal Effect (CACE) analysis (18) on the 8-week primary GAD-7 outcome to estimate intervention effects in those who completed all of the trials in a session for at least 10 out of the 12 RELAX sessions. This analysis is detailed in Section 3.4.7 below.

3.4. Statistical considerations

3.4.1. Time points

Outcomes are measured at 4, 8, and 36-weeks post-randomisation.

3.4.2. Stratification and clustering

The randomisation of individual participants is stratified by site (GSTT vs KCH vs KCL), parity (0 vs not 0 previous births at 24+ weeks) and pregnancy complications (no pregnancy complications vs pregnancy complications (current and/or past)). Therefore, we will incorporate all three stratification variables into all models as independent variables.

The data structure for the continuous outcomes is longitudinal with repeated measures over time. To account for the within-participant correlation, we will perform linear mixed effects models with all repeated measures as dependent variables, and random effects at the participant level (please refer to sections 3.2 and 3.3).

3.4.3. Missing items in scales and sub-scales

Where available we will use missing value guidance provided for scales. Where this is not available, we will prorate missing items only when there are no more than 20% missing items (i.e., for a ten-item questionnaire, prorate only where one or two items are missing) by replacing the missing item values with the mean value of the complete items for each individual. The average value for the complete items will be calculated for that individual and used to replace the missing values. The scale score will be calculated based on the complete values and these replacements. If there are more than 20% missing items, the

scale score will be considered missing. We will likely also apply this prorating procedure to the recognition test variables where applicable. Please see Appendix 2 for more detailed descriptions of instrument scoring.

3.4.4. Missing baseline data

All efforts will be made to avoid missing baseline data (i.e., requiring completion of baseline data before randomisation). However, we will summarise and report the number of participants with complete baseline measures and if this occurs, we will take an appropriate approach recommended by the literature, e.g., mean imputation as per the recommendations of White and Thompson (19).

3.4.5. Missing outcome data

Missing outcome data will be dealt with by using maximum likelihood methods to fit the mixed models described in sections 3.2 and 3.3. Maximum likelihood estimation provides valid inferences in the presence of missing observations under the assumption that all variables predicting missing outcome data are included in the models and that then the missing data mechanism is ignorable (Missing at Random, MAR).

We will therefore assess whether baseline variables predict missing outcome values so they can be included in models. First, we will create an indicator variable that will take a value of 1 if any of the primary or secondary outcomes are missing at any of the time points, and 0 if there are no missing outcome values. Then, this binary indicator will be the dependent variable of a logistic regression model with intervention arm, GAD-7 score at baseline, the three stratification variables, and weeks of gestation at baseline as independent variables. Each of the baseline demographics/participant characteristics that are listed in Section 1.8.2.2 will be added to this model background in turn as an independent variable to assess whether they predict missing outcome data. We will inspect these variables and may recode/collapse into fewer categories to avoid small categories. We will consider a variable as predictive of missing outcomes if there is a significant relationship at a 5% level. Any such variable will be included in the models for analysing the primary and secondary outcomes as described in Sections 3.2 and 3.3.

We will consider performing multiple imputation (MI) for primary and secondary outcomes only if there are post-randomisation variables that are predictive of missingness for these measures, and the proportion of participants with missing values for any of the primary or secondary outcome variables is equal or greater to 10% (20). We will employ the same methodology described above for identifying baseline predictors of missing outcome data to identify whether post-randomisation adherence to the intervention as defined in Section 2.7 predicts missing outcome data.

If multiple imputation is used, we will use a chained equations method (21), using logistic regression to estimate binary variables, multinomial regression for categorical variables of >2 levels and predictive mean matching for numerical variables. We will set the random seed number prior to starting the imputation process to ensure replicability. We will impute variables separately by intervention arm (16), where the imputation model will include adherence, baseline variables found to predict missing data, baseline and post-

randomisation measures of outcomes, the stratification factors and the pre-specified weeks of gestation baseline variable.

3.4.6. Method for handling multiple comparisons

We will make no adjustments for multiple comparisons. We will make this explicit to the reader in the final publication.

3.4.7. Method for handling non-compliance (per protocol/CACE analyses)

A per protocol analysis is not planned. However, our main secondary analysis, will be a Complier Average Causal Effect (CACE) analysis (22, 18) on the 8-week primary GAD-7 outcome to estimate intervention effects for all participants who completed all 30 trials within a session for at least 10 out of the 12 RELAX sessions.

To perform the CACE analysis, we will use an instrumental variable approach using the stata command `xtivreg` to handle the longitudinal data. We will generate a 'treated' variable coded in the intervention arm = 1 if participant completed all 30 trials within a session for at least 10 out of the 12 RELAX sessions, = 0 if they didn't, and = 0 for participants in the control arm. We will enter the intervention arm and intervention arm by time interaction term as instrumental variables for treated and treated by time interaction terms, and generally include the same covariates used for the ITT analysis (Section 3.2).

3.5. Sensitivity analyses

We will perform sensitivity analysis of the missing not at random assumption for the primary GAD-7 outcome at 8 weeks post-randomisation and secondary GAD-7 outcome at 36 weeks post-randomisation only. We will create indicator variables: 1 = missing, 0 = not missing at each time point, and then perform multiple imputation of missing data under the missing at random assumption, separately by arm (16). The imputation models will be similar to or the same as the main analysis models, including all relevant variables. We will then test alternate assumptions by adding or subtracting an amount equal to 0.2 standard deviations (SD) of the baseline GAD-7 measure to imputed data that was initially missing (23). We will entertain the following scenarios: participants missing data worsening by 0.2 SD in both arms, improving by that amount in both arms, missing in the intervention arm worsening by this amount (leaving the control arm data imputed under the missing at random assumption), and missing in the control arm improving by this amount (leaving the intervention arm data imputed under the missing at random assumption). If there are any issues with doing this analysis, we will do sensitivity analyses with the scenarios described above separately for the 8 week and 36 week time points using the Stata `rctmiss` programme (24).

3.6. Planned subgroup analyses

The study is not formally powered for subgroup analyses, but we will investigate in an exploratory fashion whether 8-week anxiety (as measured by the GAD-7) intervention effects differ by:

i. Levels of baseline anxiety (GAD-7; low = 0 to 9, moderate = 10 to 14)
ii. Pregnancy complications at baseline (yes/no). Pregnancy complications include:
    - Severe pregnancy sickness (hyperemesis gravidarum)

- High blood pressure, severe headaches, or migraines (including hypertension or preeclampsia)
- Low iron / anaemia
- Vaginal bleeding
- Placenta Previa
- Gestational diabetes
- Infections
- Sickle cell crisis
- Other

iii. Pre-existing health conditions at baseline (yes/no). Pre-existing health conditions include:

- Previous gynaecological problems
- Neurological
- Cardiovascular (e.g., heart disease)
- Respiratory (e.g., bronchitis, emphysema, asthma, COPD, pneumonia)
- Gastro-intestinal (e.g., hiatus hernia, reflux)
- Genito-urinary and renal (e.g., kidney disease, UTIs)
- Endocrine (e.g., diabetes, thyroid disease)
- Musculo-skeletal (e.g., osteoarthritis, rheumatoid arthritis)
- Cancer
- Other

The two-way trial arm by time point interaction terms will be extended to three-way trial arm by time point by baseline subgroup variable terms that will be added to the main analysis model for the primary anxiety outcome. If these interaction terms are statistically significant with respect to the subgroup variable of interest, we will present intervention versus control estimates at 8-weeks post-randomisation by baseline anxiety, pregnancy complications, and/or pre-existing conditions levels.

This analysis will be clearly stated as exploratory in the primary paper/report and will be interpreted accordingly.

3.7. Exploratory analyses

No exploratory analyses are planned.

3.8. Interim analyses

No interim analyses are planned for this study.

4. Software

4.1. Data management

Data extracts will be obtained by the trial statistician prior to each DMEC meeting and at the end of the study. Extracts will be requested from Avegen via email and will be received by the trial statistician within three days of request. All data including baseline and outcome measures will be stored and sent directly from the Avegen platform.

4.2. Statistical software

Besides data collection, storage and extraction, we will perform all data cleaning, processing, analysis, and reporting using Stata version 18 or higher. If required, we will consider the use of other suitable statistical packages such as R or Mplus.

5. References

1. Spitzer RL, Kroenke K, Williams JBW, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. Arch Intern Med. 2006 May 22;166(10):1092–7.

2. Mathews A, Mackintosh B. Induced emotional interpretation bias and anxiety. J Abnorm Psychol. 2000 Nov;109(4):602–15.

3. Hirsch CR, Meeten F, Gordon C, Newby JM, Bick D, Moulds ML. Repetitive Negative Thinking and Interpretation Bias in Pregnancy. Clin Psychol Eur [Internet]. 2020 Dec 1 [cited 2023 Jan 17];2(4). Available from: https://cpe.psychopen.eu/index.php/cpe/article/view/3615

4. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med. 2001 Sep;16(9):606–13.

5. McEvoy PM, Thibodeau MA, Asmundson GJG. Trait Repetitive Negative Thinking: A Brief Transdiagnostic Assessment. J Exp Psychopathol. 2014 Nov 1;5(3):1–17.

6. Meyer TJ, Miller ML, Metzger RL, Borkovec TD. Development and validation of the Penn State Worry Questionnaire. Behav Res Ther. 1990;28(6):487–95.

7. Cox JL, Holden JM, Sagovsky R. Detection of postnatal depression. Development of the 10-item Edinburgh Postnatal Depression Scale. Br J Psychiatry J Ment Sci. 1987 Jun;150:782–6.

8. Somerville S, Dedman K, Hagan R, Oxnam E, Wettinger M, Byrne S, et al. The Perinatal Anxiety Screening Scale: development and preliminary validation. Arch Womens Ment Health. 2014 Oct;17(5):443–54.

9. Mundt JC, Marks IM, Shear MK, Greist JH. The Work and Social Adjustment Scale: a simple measure of impairment in functioning. Br J Psychiatry J Ment Sci. 2002 May;180:461–4.

10. Jacobson NS, Truax P. Clinical significance: a statistical approach to defining meaningful change in psychotherapy research. J Consult Clin Psychol. 1991 Feb;59(1):12–9.

11. Toussaint A, Hüsing P, Gumz A, Wingenfeld K, Härter M, Schramm E, et al. Sensitivity to change and minimal clinically important difference of the 7-item Generalized Anxiety Disorder Questionnaire (GAD-7). J Affect Disord. 2020 Mar 15;265:395–401.

12. Hirsch CR, Krahé C, Whyte J, Krzyzanowski H, Meeten F, Norton S, et al. Internet-delivered interpretation training reduces worry and anxiety in individuals with generalized anxiety disorder: A randomized controlled experiment. J Consult Clin Psychol. 2021;89:575–89.

13. D. Machin. Sample Size Tables for Clinical Studies. Wiley; 1997.

14. van Ballegooijen W, Cuijpers P, van Straten A, Karyotaki E, Andersson G, Smit JH, et al. Adherence to Internet-Based and Face-to-Face Cognitive Behavioural Therapy for Depression: A Meta-Analysis. García AV, editor. PLoS ONE. 2014 Jul 16;9(7):e100674.

15. Schulz KF, Altman DG, Moher D, CONSORT Group. CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. PLoS Med. 2010 Mar 24;7(3):e1000251.

16. Sullivan TR, White IR, Salter AB, Ryan P, Lee KJ. Should multiple imputation be the method of choice for handling missing data in randomized trials? Stat Methods Med Res. 2018 Sep 1;27(9):2610–26.

17. Fritz M, Taylor A, MacKinnon D. Explanation of Two Anomalous Results in Statistical Mediation Analysis. Multivar Behav Res. 2012 Feb 8;47:61–87.

18. Dunn G, Emsley R, Liu H, Landau S, Green J, White I, et al. Evaluation and validation of social and psychological markers in randomised trials of complex interventions in mental health: a methodological research programme. Health Technol Assess. 2015 Nov;19(93):1–116.

19. White IR, Thompson SG. Adjusting for partially missing baseline measurements in randomized trials. Stat Med. 2005 Apr 15;24(7):993–1007.

20. Buuren S van. Flexible imputation of missing data [Internet]. Second edition. Boca Raton, FL: CRC Press; 2018 [cited 2023 Feb 7]. 415 p. (Interdisciplinary statistics). Available from: https://search.ebscohost.com/login.aspx?direct=true&scope=site&db=nlebk&db=nlabk&AN=1893427

21. White IR, Royston P, Wood AM. Multiple imputation using chained equations: Issues and guidance for practice. Stat Med. 2011;30(4):377–99.

22. Angrist JD, Imbens GW, Rubin DB. Identification of Causal Effects Using Instrumental Variables. J Am Stat Assoc. 1996 Jun 1;91(434):444–55.

23. White IR, Horton NJ, Carpenter J, Statistics R in M and S, Pocock SJ. Strategy for intention to treat analysis in randomised trials with missing outcome data. BMJ. 2011 Feb 7;342:d40.

24. White IR, Carpenter J, Horton NJ. A mean score method for sensitivity analysis to departures from the missing at random assumption in randomised trials. Stat Sin. 2018 Oct;28(4):1985-2003.

6. Appendix 1 Schedule of Assessments

#	Measure	Screening assessment	T0: Pre-intervention baseline assessment 16-28 weeks gestation	T1: 4 weeks post-randomisation 20-32 weeks gestation	T2: 8-weeks post-randomisation 24-36 weeks gestation	T3: 36-weeks post-randomisation 12-24 weeks post-birth
Avegen database						
1	Eligibility Questions	X				
2	Informed Consent		X			
3	Demographics	X	X			
4	Psychological Contact and Treatment		X	X	X	X
5	Anxiety (GAD-7)	X	X	X	X	X
6	Interpretation bias (RT)		X	X		
7	Depression (PHQ-9)	X	X	X	X	X
8	Repetitive Negative Thinking (RTQ-10 trait)	X	X	X	X	X
9	Worry (PSWQ)		X	X	X	X
10	Perinatal depression (EPDS)		X	X	X	X
11	Perinatal anxiety (PASS)		X	X	X	X
12	Work and Social Functioning (WSAS)		X	X	X	X
13	Changes to medication, contact with a health professional.			X	X	X
14	Pregnancy complications	X		X	X	
15	Birth complications and non-routine health professional contact			X	X	X
16	Adverse events			X	X	X
17	Adherence	Ongoing through 4-weeks of intervention (intervention arm only)				

7. Appendix 2 Brief description of instrument scoring

- GAD-7 (2) is a well validated and widely used 7 item measure of anxiety. Each item is rated for reoccurrence, ranged from 0 (not at all) to 3 (nearly every day), and a total score ranges from 0 to 21. Scores of 5, 10, and 15 are taken as the cut-off points for mild, moderate, and severe anxiety, respectively. When used as a screening tool, further evaluation is recommended when the score is 10 or greater.
- PHQ-9 (3) is a well validated 9 item questionnaire used to measure symptoms of depression. Each item is rated for reoccurrence, ranged from 0 (not at all) to 3 (nearly every day), and a total score ranges from 0 to 27. Depression Severity: 0-4 none, 5-9 mild, 10-14 moderate, 15-19 moderately severe, 20-27 severe.
- RTQ-10 (trait) (4) is a validated transdiagnostic measure of RNT. Each of the ten items, scored from 1 to 5, with 1 representing "not true at all", 3 representing "somewhat true" and 5 representing "very true" are summed to produce a total score between 10 to 50. Higher scores indicate greater level of repetitive negative thinking.
- The PSWQ (6) is an established measure of trait worry (i.e., general tendency to worry). This measure is included because worry is the primary form of RNT in relation to anxiety. The 16 items, ranging from 1 (not at all typical of me) to 5 (very typical of me), are summed to produce a total score between 16 to 80. Items 1, 3, 8, 10 and 11 are 'reverse scored' e.g. if 1 is selected for one of these items, the item is scored as 5 and if 5 is selected for one of these items, the item is scored as 1. Higher score indicate greater levels of worry.
- EPDS is the most common screening tool for perinatal depression. The ten items are scored from 0 to 3 and summed to produce a total score between 0- 30. Items 3, 5, 6, 7, 8, 9 and 10 are 'reverse scored'. Higher score indicates more severe perinatal depression, (7).
- The PASS (8) measures perinatal specific anxiety. Scores for each of the 31 items ranging from 0 to 3 are summed to produce a total score between 0-93. Higher score indicates greater levels of perinatal anxiety. We will not analyse separate subscale scores in the RELAX study.
- The Work and Social Adjustment Scale (13) is a 5-item patient self-report measure, which assesses the impact of a person's mental health difficulties on their ability to function in terms of work, home management, social leisure, private leisure and personal or family relationships. Each item is rated between 0 (no impairment) and 8 (very severe impairment), and a summed total score ranges from 0 to 40. If participants are not working, the item related to work will not be applicable. In this case, the WSAS score will be prorated: the mean of available scores will be imputed for the NA item, and the total score calculated.

8. Appendix 3 Manuscript Deviations from or Omissions relative to the Statistical Analysis Plan

• Baseline measures that were not measured across all participants are not listed in the Baseline measures table (e.g. type of pre-existing physical health conditions).

• Currently receiving psychological treatment (Y/N) at T1/T2/T3.

• Currently taking medication for psychological disorders (Y/N) at T1/T2/T3.

• RELAX usage variables; number of sessions completed, number of trials completed per session, mean comprehension question accuracy for each session, positivity and vividness scores for each session.

• Proportion who withdrew from analysis (number presented in manuscript).