Statistical Analysis Plan

Cluster Randomised Controlled Trial of a Service to Support the Mental Health and Coping of Parents with Excessively Crying Infants (Surviving Crying)

SAP Version: 1.0 Final
Date: 14-04-2026

Based on protocol/protocol amendment: Brown_19_432_Surviving_Crying
Version: 10.0
Date: 16-10-2025
Trial Registration: ISRCTN 73761296

Sponsor: Nottingham Trent University
50 Shakespeare Street, Nottingham, NG1 4FQ

Trial Statistician: Ana Suazo Di Paola
University of Leicester, Maurice Shock Building,
University Road, Leicester, LE1 7RH

Chief Investigator: Dr Anne Felton
The Institute of Health and Allied Professions
Clifton Lane, Clifton, Nottingham, NG11 8NS

Trial Manager: Yvette Walters
University of Leicester, Maurice Shock Building,
University Road, Leicester, LE1 7RH

PRINTED COPY MAY NOT BE THE MOST UP TO DATE VERSION

Revision History

Version	Date	Author(s) and Role	Summary of Changes/Comments
1.0	14-04-2026	Ana Suazo Di Paola, Trial Statistician	Final version

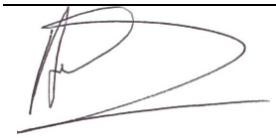
SAP Approval for Finalized Version:

Trial Statistician **Ana Suazo Di Paola**



Signature 14/04/2026 **Date**

Chief Investigator **Dr Anne Felton**



Signature **14/04/2025** **Date**

Senior Statistician **Cassey Brookes**



Signature 14/04/2026 **Date**

PRINTED COPY MAY NOT BE THE MOST UP TO DATE VERSION

List of Abbreviations

AE	Adverse Event
ANCOVA	Analysis of Covariance
BAME	Black, Asian and Minority Ethnic
CBT	Cognitive Behavioural Therapy
CONSORT	Consolidated Standards of Reporting Trials
CRF	Case Report Form
DSMC	Data Safety Monitoring Committee
e-CRF	Electronic case report form
EPDS	Edinburgh Postnatal Depression Scale
EQ5D	Quality of Life Questionnaire
GAD-7	Generalised Anxiety Disorder Assessment
GCP	Good Clinical Practice
GDPR	General Data Protection Regulation
GEE	Generalised Estimating Equations
GLM	Generalised linear Models
HRA	Health Research Authority
HCP	Health Care Professional
HRQoL	Health Related Quality of Life
HV(s)	Health Visitor(s)
ICC	Intraclass Correlation Coefficient
ICH GCP	International Conference on Harmonisation Good Clinical Practice
iHV	Institute of Health Visiting
ISRCTN	International Standard Randomised Controlled Trial Number
LCTU	Leicester Clinical Trials Unit
MCID	Minimally Clinically Important Difference
MCQ	Maternal Confidence Scale
NICE	The National Institute for Health and Care Excellence
NIHR	National Institute for Health Research
NIHR HTA	National Institute for Health Research Health Technology Assessment
PIS	Participant Information Sheet
RCT	Randomised Controlled Trial

PRINTED COPY MAY NOT BE THE MOST UP TO DATE VERSION

RDC	Remote Data Capture
REC	Research Ethics Committee
R&I	Research and Innovation
SAE	Serious Adverse Event
SAP	Statistical Analysis Plan
SC	Surviving Crying
SD	Standard Deviation
SOP	Standard Operating Procedure
TMF	Trial Master File
TSC	Trial Steering Committee
UKCRC	UK Clinical Research Collaboration
UoL	University of Leicester

PRINTED COPY MAY NOT BE THE MOST UP TO DATE VERSION

Contents

1 Introduction 6
1.1 Background and rationale 7
1.2 Aim 8
1.3 Study Objectives 8
1.3.1 Primary Objective 8
1.3.2 Secondary Objectives 8
1.3.3 Exploratory objectives 8
1.3.4 Subgroup Objectives 8
1.4 Study Design 9
1.4.1 Participants 9
1.4.2 Intervention Arm (SC Package) 10
1.4.3 Control Arm 11
1.4.4 Sample Size 11
1.4.5 Randomisation and Blinding 12
1.5 Visit Schedule 13

2 Quality Control and Data Validation 15
2.1 Derived/ Computed Variables 15
2.2 Validation of the Analyses 16
2.3 Analysis Software 16

3 Interim Analysis and Data Review 16
3.1 Interim Analysis 16
3.1.1 Internal Pilot – Primary Objective 17
3.2 Data Review 17

4 Descriptive Analyses 17
4.1 Disposition 17
4.2 Demographic and Baseline Characteristics 17
4.3 Reporting of Losses to Follow-up and Withdrawals of Consent 18
4.4 Treatment Compliance 18
4.5 Other Variables 18

5 Protocol Deviations 19
5.1 Major Deviations 19
5.2 Minor Deviations 19

6 Definition Of Population for Analyses 20
6.1 Complete-Case Population 20
6.2 Intention-to-Treat Population 20
6.3 CACE Population 20
6.4 Safety Population 20

7 Outcome Measures 21
7.1 Primary Outcome 21
7.1.1 Definition and Derivation of Primary Outcome 21
7.1.2 Hypothesis to be investigated 21
7.2 Secondary Outcomes 21
7.2.1 Definition and Derivation of Secondary Outcomes 21
7.2.2 Hypotheses to be investigated 23
7.3 Exploratory Outcomes 23

PRINTED COPY MAY NOT BE THE MOST UP TO DATE VERSION

7.4 Health Economic Outcomes 23

8 Primary Outcome Analysis 24
8.1 Primary Analysis of Primary Outcome 24
8.1.1 Primary Estimand 24
8.1.2 Statistical Model 25
8.2 Secondary Analyses of Primary Outcome 25
8.2.1 Intention to Treat Analysis 26
8.2.2 CACE Analysis 27
8.2.3 EPDS scores out of timeframe Analysis 27
8.3 Adjustment of P values for Multiple Testing 28
8.4 Missing Data 28
8.5 Pre-specified Subgroup Analysis 28
8.6 Sensitivity Analysis 29

9 Secondary Outcome Analyses 29
9.1 Primary Analysis of Secondary Outcomes 29
9.2 Secondary Analyses of Secondary Outcomes 32
9.3 Missing Data 32
9.4 Pre-specified Subgroup Analysis 32

10 Additional Analyses 32
10.1 Exploratory Analyses of the Primary Outcome 32

11 Safety and Adverse Events Reporting 33

12 References 34

13 Appendices 35
13.1 Appendix 1: Templates for Tables, Listings and Figures 35

PRINTED COPY MAY NOT BE THE MOST UP TO DATE VERSION

1 Introduction

This Statistical Analysis Plan (SAP) describes the planned analysis and reporting for the main paper(s) results from the “Cluster Randomised Controlled Trial of a Service to Support the Mental Health and Coping of Parents with Excessively Crying Infants (Surviving Crying)”. The structure and content of this SAP follows the LCTU quality assured standard operating procedures for statistical analysis which is based on the guidelines issued by the International Conference on Harmonization (ICH) topic E9 on Statistical Principles in Clinical Trials. The plan of the statistical analysis should be detailed here taking into account the protocol and related amendments. Subsequent/post-hoc analyses of a more exploratory nature will not be bound by this SAP, though they are expected to follow the same broad principles laid down here.

The SAP will be amended if there are substantial changes to the planned analyses, and in any case will be finalized before the database lock for this study. Any deviations from the SAP will be described and justified in the final report of the trial.

The analysis strategy will be available on request when the principal papers are submitted for publication in a journal. Suggestions for subsequent analyses by journal editors or referees, will be considered carefully, and carried out as far as possible in line with principles of this SAP.

The reader of this SAP is encouraged also to read the clinical trial protocol.

Throughout the document: Any verbatim text from the protocol is provided inside a box:

Text from the protocol

PRINTED COPY MAY NOT BE THE MOST UP TO DATE VERSION

1.1 Background and rationale

Parents report that around 20% of infants cry persistently without an apparent reason during the first four postnatal months. There is objective evidence that they are often correct in reporting prolonged crying. Based on 2018 live birth rates, this affects some 146,000 UK families per year. Traditionally, this phenomenon was known as 'colic', implying that the crying is due to gastro-intestinal disorder and pain. In spite of many studies, however, the crying remains unexplained and there is no reliable way to prevent or resolve it. Recently, the need for a focus on the impact of this crying on parent/carers, and subsequent outcomes, has become apparent. Crying in early infancy includes long lasting bouts and is sometimes resistant to normal soothing measures, which parents find particularly challenging. They worry that the crying is a sign their infant is unwell. However, the impact on parental emotions and actions depends partly on how parents cope with the crying. Anxiety and emotional arousal influence how parents interpret infant crying and this can precipitate parental depression. Other adverse outcomes include premature termination of breastfeeding, over-feeding, poor parent-child relationships and child development outcomes, and infant abuse in rare cases.

Against this background, the lack of evidence-based NHS services to support parents in managing infant crying is striking. Instead, parents turn to popular media, which give conflicting advice, or take infants to GPs or hospitals, adding to NHS costs, which are substantial. In Canada, a support programme for parents reduced hospital admissions for infant abusive head trauma and proved to be cost-effective over the long-term. Based on these findings, the 'Surviving Crying' (SC) feasibility study was designed to develop materials which support UK parents who judge their infant to be crying excessively. The resulting package includes a website, information booklet, and short Cognitive Behaviour Therapy (CBT) programme tailored to support parents' coping with excessive infant crying. Following NICE guidelines, the website or booklet were intended to be inexpensive and available to all parents; the CBT sessions were intended for families with greater need. To evaluate the package provisionally, 52 parents with excessively crying infants were referred to the study by HVs and completed standardised measures before receiving the support package (baseline) and afterwards. Almost all parents accessed the website or booklet; half accessed the CBT sessions. At baseline, 30 of 52 parents met clinical criteria for depression, halving to 15 after receiving the SC package. In regression analyses, these improvements were not explained by changes in infant crying. Reductions occurred in the number of parents reporting the crying to be a large or severe problem (from 28 to 3 parents) or feeling very or extremely frustrated by the crying (from 31 to 1 parent). Other findings included improvements in parents' confidence, knowledge of infant crying, sleep, and reductions in NHS contacts. All 52 parents and 85% of 50 HVs agreed the package should be included in the NHS and 94% of the HVs wanted materials of this kind to be included in HV training. Most (86%) of the parents were willing to take part in a controlled trial to fully evaluate the SC package.

Because HVs and allied professionals provide primary healthcare support for all UK parents and infants, the research so far has been carried out together with HVs and the SC package was designed to be delivered by HVs and compatible with statutory HV 'universal' and 'universal-plus' services. There is also robust evidence that nurses can deliver tailored CBT sessions similar to those in the SC package.

PRINTED COPY MAY NOT BE THE MOST UP TO DATE VERSION

Unfortunately, HV services have been subject to recent cuts and re-organisations which query whether HVs can deliver the SC materials in practice. In the feasibility study, researchers delivered the website and booklet; a psychologist the CBT sessions. Potentially, HVs might deliver digital services such as the SC website, other professionals provide CBT. We continue to view health visiting as the best way to deliver these supports as a joined-up primary healthcare service, but accept that has to be demonstrated. It also needs to be demonstrated that the reductions in depression symptoms found in the SC feasibility study are due to the SC intervention, rather than to care as usual. To provide the evidence needed, we have designed a trial to compare SC with usual services for their effectiveness and cost.

1.2 Aim

The overall aim of the trial is to test the clinical effectiveness and cost effectiveness of statutory health visiting (HV) services supplemented by SC-based service compared to statutory health visiting (HV) services alone.

The aim corresponding to the cost effectiveness is not included in this SAP as it will be addressed in the Health Economics Analysis Plan.

1.3 Study Objectives

1.3.1 Primary Objective

To measure the effect of the SC intervention compared to usual care in participating parents as measured with the EPDS at 8 weeks follow-up.

1.3.2 Secondary Objectives

To measure the effect of the SC intervention compared to usual care in participating parents in terms of secondary outcome measures.

The secondary objective corresponding to the Health Economics analysis is not included in this SAP as it will be addressed in the Health Economics Analysis Plan.

1.3.3 Exploratory objectives

Exploratory analyses of the primary outcome will investigate the effect of differing levels of compliance over and above the minimum level

1.3.4 Subgroup Objectives

Subgroup analyses of the primary outcome will be performed on the following factors: SC HV accredited in CBT, social deprivation index and ethnicity.

CBT accreditation: SC Health Visitors will be classed as having received CBT accreditation by using CBT session data on ‘SC Trained HV obtained accreditation (Y/N)’ and ‘date accreditation was awarded’. It is important to note that the classification of HV

PRINTED COPY MAY NOT BE THE MOST UP TO DATE VERSION

teams into accredited vs non-accredited will be used to conduct an exploratory analysis of the primary outcome.

Social deprivation index: the quintiles approach (using national cut-points) will be implemented to classify each participant into one of 5 groups (20% each), with quintile 5 representing the "most deprived" group and quintile 1 representing the "least deprived" group (20%). The local research team determines each participant's social deprivation score using the participant's postcode. The score is collected in the clinical trial's database.

Ethnicity: ethnicity data will be used to classify participants into two categories: white or other.

Level of Education: highest level of education data collected for each participant at Baseline will be used to map each participant's education level to the UK's 6-level system in accordance with the ISCED and to classify participants into two categories: the first category will include ISCED levels 0-3 and the second one will include levels 5-7. This classification will be done for the purposes of conducting an exploratory subgroup analysis.

1.4 Study Design

Cluster Randomised Controlled Trial following CONSORT and Ottawa guidelines for such trials.
The proposed trial is in two phases: the pilot and main phase. The trial will run an internal pilot to prove recruitment of the relevant number of participants during 9 months from approximately 20 clusters.
Agreed Stop-go criteria:

- Randomise 20 Clusters (HV Teams) from 2-4 sites.
- Train and certify 10 SC Trained HVs to deliver the Surviving Crying CBT based sessions to parents.
- Achieve participant recruitment (target 56 participants) at an average rate of 0.4/cluster/month.

Progression:
GO: 100% of target achieved
REVISE: At least one target falls below 100% but achieves at least 80%

STOP: At least one target falls below 80%

1.4.1 Participants

Parents/carers aged 18 years or older with an infant up to 20 weeks old at time of consent who express concern about their infant's excessive crying will be deemed as eligible participants. However, parents/carers with any seriously ill infant that requires inpatient hospitalisation or parents/carers involved with any other clinical trial at time of consent or 3 months prior will be deemed as not eligible to take part in the trial.

PRINTED COPY MAY NOT BE THE MOST UP TO DATE VERSION

1.4.2 Intervention Arm (SC Package)

1. SC Booklet
2. SC Website
3. Cognitive Behaviour Therapy (CBT) Session's

Participants will be given the choice as to which of the three possible intervention packages they can access; the SC website, a copy of the SC booklet and CBT sessions. They can chose all of these elements or a combination of any of the three. The research administrator will provide the participant with individualised access to SC website, post or email the SC booklet if requested and notify the SC Trained HV if the participants requests CBT sessions.
The SC Trained HV will organise the CBT sessions with the participants where requested. Participants will also have access to usual NHS HV services delivered by their regular HV.
Where randomised to the intervention arm, participants should receive delivery of the first session within 3 working days of the date the SC Trained HV was notified of their request. Continuity between members of the research team and intervention delivery and responding quickly once a participant is enrolled should be ensured. Participants will be contacted by telephone text or email. This is to help retention in the trial and ensure support is received quickly by the participant who has requested this.

Participants will normally be offered between 1 and 4 sessions of one to one support from a SC Trained Health Visitor which would normally be delivered at home but could be delivered remotely if circumstances (such as Covid-19) demand this. Additional sessions up to a maximum of 6 can be provided where the SC CBT trained HV considers these to be needed. All sessions will be delivered on a weekly basis but must not exceed 6 weeks from date of the first CBT session.

Participants sessions will be recorded (if selected) to allow SC Trained Health Visitors to be monitored and assessed to certify their competence in level 2 CBT skills.
SC Trained HV will audio record the sessions with parents, using an access restricted trial mobile phone/laptop, no identifiable information will be recorded.

All SC trained HV's will upload the recording using trial ID numbers only, to the NIHR Hub secure site where restricted access will only be provided to the SC CBT practitioner trainer. Once accreditation has been received no further recordings will be required and all recordings will be destroyed. We will ask participants and SC Trained HVs to consent to the recordings of these sessions.

Participants will also have access to usual NHS HV services delivered by their regular HV throughout the duration of the study.

Website use will be measured by the objective metrics provided by Google Analytics.

SC Trained HV's will record on a paper CRF the type and amount of intervention each participant in the intervention received. This data will be entered into the SC Database.
This will help to assess cost of their overall package to be compared to routine services.

PRINTED COPY MAY NOT BE THE MOST UP TO DATE VERSION

1.4.3 Control Arm

Participants allocated to the control arm, will continue to receive usual NHS HV services. However they can request on completion of the trial (16 weeks post baseline assessment, the SC booklet and/or access to the SC website.
They will be provided with an individualised access code for the website (identifiable as control participant).

1.4.4 Sample Size

The sample size has been derived to detect a 2-point difference in the mean EPDS score (10.5 for the control group and 8.5 for the intervention group). With a 5% level of significance (two-sided), 90% power and assuming a Gamma family distribution with a log link function and scale parameters (k) 3.78 and 3.06 for the control and intervention groups respectively (1:1 allocation), equation 14 in 37 estimates that a total of 280 participants are required for an individually randomised trial.

Calculating the sample size required for a cluster randomised trial involves inflating the estimated sample size of 280 assuming individual randomisation by a design effect.
Assuming an average cluster size of 4 with data available for the primary analysis, with a range of 2 to 12 giving coefficient of variation equal to 0.625 and ICC of 0.037 (observed in 40 results in a design effect of 1.17 (using equation 18 in [3]). Inflating 280 by 1.17 gives a target sample size of 328 (164 per arm). Furthermore, a sample size of 328 and 4 participants per cluster would require a total of 82 clusters to be randomised (41 per arm). Standard convention means we add 1 cluster per arm to allow for the use of the t distribution and 1 extra cluster per arm is also added to account for potential whole cluster drop out. The sample size is further inflated by 12% to allow for individual loss of data and non-compliance with the primary outcome (data loss occurred in 9% of participants in our feasibility study [24], but we chose a more conservative rate). Therefore, each cluster should recruit 4.55 participants on average to allow for drop out (our projections show clusters could recruit an average of 4.8 so this is feasible). This gives a total target sample size of 86 clusters (43 per arm) and 392 participants (196 per arm). The effect on the sample size of assuming alternative model parameters, MCIDs, ICC values (0.01 and 0.05) and wider cluster size ranges was evaluated. These sensitivity analyses showed that the trial is robust to differences in these parameters.

In light of the increased variability that was observed within the first quarter of 2024 in terms of average cluster characteristics (i.e. mean, minimum and maximum value), with some HV teams (clusters) recruiting a minimum of 1 participant at some sites and one cluster at Newham site recruiting a maximum of 21 participants, the TMG and the LCTU made the decision to close the latter cluster to recruitment and to cap the cluster size across all sites at 15. The implementation of this approach will help ensure that the variability in cluster size is minimised and remains consistent across all sites and that the target sample size is achieved within the approved study timelines without requiring the addition of more clusters and/or participants in order to avoid being statistically underpowered.

PRINTED COPY MAY NOT BE THE MOST UP TO DATE VERSION

1.4.5 Randomisation and Blinding

Randomisation will be managed and carried out by LCTU. To ensure HV team clusters are evenly distributed, they will be randomly allocated either to the control or intervention group on a 1:1 basis, stratified by site.

Allocation of each HV Team cluster will be carried out when the first parent/carer's informed consent is obtained in each HV team.

The regular HV will be not be notified which group their HV team has been allocated at any point during the trial.

The researcher who has performed the consent with the parent/carer and subsequently completes all assessments, will also be blinded to the allocation of each cluster and will not know which allocation an individual participant belongs to.

Any unintended unblinding by the researcher will be recorded and another assessor will complete all further assessments for that participant. At each assessment, the outcomes assessor will be asked to guess the treatment allocation of the participant.

There will be a Research Administrator team member who will remain unblinded to the allocation and will be informed by LCTU of which of the two trial arms each HV team cluster has been allocated to and which HV team cluster subsequent participants belong to. This Research Administrator team member will be responsible for informing all participants after they have attended their baseline assessment, of their arm allocation within the trial. They will also be informed that they should not disclose the results of this allocation to the researcher who is performing the follow-up visits or their regular HV. SC trained HV will be informed if a new participant is enrolled into an intervention cluster and has requested CBT to enable delivery of the intervention. The SC Trained HV whilst not openly unblinded to cluster allocations, will be aware of those participants that require them to conduct intervention delivery.

PRINTED COPY MAY NOT BE THE MOST UP TO DATE VERSION

1.5 Visit Schedule

Outcome measures	**Baseline Measures. (Consent to study)**	**8** Week (From Baseline Assessment) **Follow up Measures**	**16 Week** (From Baseline Assessment) **Follow Up Measure**
Visit Window	(Within 3 working days of identification of parent/carer)	(+/- 1 week within follow up due date)	(+/- 1 week within follow up due date)
EPDS	✓	✓	✓
GAD7	✓	✓	✓
EQ-5D-5L	✓	✓	✓
Parenting Confidence Questionnaire (Maternal Confidence Questionnaire)*	✓	✓	✓
Couple Satisfaction Index (CSI-4)	✓	✓	✓
Health Economics	✓	✓	✓
Crying Knowledge Scale **	✓	✓	
Parents/carers' ratings of how frustrating the crying is for them	✓	✓	✓
Parents/carers' ratings of how problematic the crying is for them	✓	✓	✓
Parental Sleep Adequacy***	✓	✓	✓
Social Support ****	✓	✓	✓
Health, breast-feeding maintenance	✓	✓	✓
Parent/HV Evaluation of Services received.		✓	

PRINTED COPY MAY NOT BE THE MOST UP TO DATE VERSION

*. (Originally Maternal Confidence Q. replacing ‘mothers’ with ‘parents’).
*** 8 items from Barr RG, Barr M, Fujiwara T, et al. CMAJ 2009;180. DOI:10.1503/cmaj.081419.
*** 2 items from Pittsburgh Sleep Quality Index, following Hiscock et al. 2014
**** 2 items from Hiscock et al. 2014

PRINTED COPY MAY NOT BE THE MOST UP TO DATE VERSION

2 Quality Control and Data Validation

2.1 Derived/ Computed Variables

Rules for calculation of derived variables, including definitions that will be programmed from the data (other than those specified in the primary and secondary outcome measures section below) along with data transformations and methods for handling multiple observations will be included in this section.

Sleep of parents/carers over time

The average hours of sleep per night and the average hours of sleep per 24 hours (including day-time sleeping) during the last 2 weeks prior to the Baseline, 8 weeks and 16 weeks follow-up visit will be used to calculate the total hours of sleep (night time) and the total hours of sleep (per 24 hours) during the last 2 weeks prior to the study visits.

Total hours of sleep (night time) during last 2 weeks = Average hours of sleep per night during last 2 weeks * 14

Total hours of sleep (including day-time sleeping) = Average hours of sleep per 24 hours * 14

HV time until CBT training accreditation

The HV time until accreditation is defined as the time, measured in months, from the date of initial training, date of supervision or date of refresher training (whichever is available and/or first) until the date the HV received CBT accreditation. In the event that any of these dates is before the date the HV recruited their first participant, the date of recruitment of the HV's first participant will be used instead.

HV time until CBT training accreditation = (Date of Accreditation – Date of initial training/date of supervision/date of refresher training/date of recruitment of first participant)/30.4

CBT sessions delivered by accredited/unaccredited HVs

Whether individuals received CBT by accredited HVs will be summarised using two different definitions:

1) **Accredited/Unaccredited at time of delivery of sessions:** a participant is considered to have received accredited CBT sessions if all of their CBT sessions were delivered by a HV accredited prior to the delivery of the first session. A participant is considered to have received unaccredited CBT sessions if any of their CBT sessions were delivered by a HV prior to their accreditation date.
2) **Accredited/Unaccredited at some point during trial:** a participant is considered to have received accredited CBT sessions if their CBT sessions were delivered by a HV that received accreditation during the course of the trial. A participant is considered to have received unaccredited CBT sessions if their CBT sessions were delivered by a HV that never received accreditation during the course of the trial.

Timing of follow-up assessments

The time until the 8 and 16 weeks follow-up assessments is defined as the time, measured in weeks, from Baseline assessment date until the date the follow-up assessments were carried out.

Time until 8 weeks visit = (8 weeks assessment date – Baseline assessment date)/7

Time until 16 weeks visit = (16 weeks assessment date – Baseline assessment date)/7

PRINTED COPY MAY NOT BE THE MOST UP TO DATE VERSION

2.2 Validation of the Analyses

A statistician, other than the nominated trial statistician, will validate the analyses listed below. When non-validated statistical software packages are used these validations will be carried out using different statistical software, if possible. Any unresolved discrepancies will be reported in the statistical report.

Primary Outcome:

Edinburgh Postnatal Depression Scale (EPDS) score at 8 weeks.

Secondary Outcomes:

• Edinburgh Postnatal Depression Scale (EPDS) score 16 weeks post baseline as a continuous outcome in addition to the proportion of parents/carers with EPDS score of 10 or more at 16 weeks post baseline.

2.3 Analysis Software

All analysis will be carried out using appropriate validated statistical software such as STATA, SAS or R statistical software. The relevant package and version number will be recorded in the statistical report.

3 Interim Analysis and Data Review

3.1 Interim Analysis

The proposed trial is in two phases: the pilot and main phase. The trial will run an internal pilot to prove recruitment of the relevant number of participants during 9 months from approximately 20 clusters.

Agreed Stop-go criteria:
- Randomise 20 Clusters (HV Teams) from 2-4 sites.
- Train and certify 10 SC Trained HVs to deliver the Surviving Crying CBT based sessions to parents.
- Achieve participant recruitment (target 56 participants) at an average rate of 0.4/cluster/month.

Progression:
GO: 100% of target achieved
REVISE: At least one target falls below 100% but achieves at least 80%
STOP: At least one target falls below 80%

A progress summary of the trial will be prepared and presented to the Data Monitoring Committee (DMC) during the DMC meeting to be held in May 2023 (end of pilot phase). The following primary outcomes of the internal pilot will be reported:

- The total number of clusters (HV teams) randomised from 2-4 sites.
- The total number of parents/carers recruited into the trial.

PRINTED COPY MAY NOT BE THE MOST UP TO DATE VERSION

3.1.1 Internal Pilot – Primary Objective

To assess ability to recruit HV teams, train HVs in the delivery of the SC intervention, and recruit parents of excessively crying infants in a 9-month internal pilot study.

3.2 Data Review

The Surviving Crying trial has a Data Monitoring Committee (DMC) which is composed of three independent members: a chairperson who also acts as the independent statistician, a clinician experienced in the area and another clinician experienced in research.

The remit of the DMC involves being responsible or safeguarding the interest of study participants, assessing the safety of the interventions during the study and for monitoring the overall conduct of the study.

In accordance with the DMC Charter v1.0, the DMC will meet approximately every 12 months, with the first meeting taking place in June 2022. It is important to note that after the initial meeting, the DMC have been meeting approximately every 6 months.

4 Descriptive Analyses

4.1 Disposition

Trial disposition will be presented on the ITT population with respect to both clusters and participants. The number of clusters randomised, number of clusters withdrawn, number of clusters providing at least one primary outcome observation and cluster size will be summarised by treatment arm and overall. The number of participants registered, withdrawn consent/lost, provided primary outcome data, completed trial follow-up, intervention compliance, protocol deviations, length of stay in the trial and blinding status will be summarised by treatment arm and overall. All reasons for withdrawal and non-completion will be listed where available.

A CONSORT diagram will display the flow of clusters and participants through the trial including the eligibility, reasons for exclusion, number of clusters randomised to each treatment arm, number of participants registered, number of participants who completed/did not complete their assessment at each of the follow-up time points, number of participants who were lost to follow-up and number of participants analysed in each population.

A graph of cumulative recruitment will be presented in addition to summaries of recruitment (e.g. start and end date) by site and overall.

Data completeness (i.e. CRF return rate) will also be summarised by treatment arm and overall for each CRF.

4.2 Demographic and Baseline Characteristics

Numbers (with percentages) for binary and categorical variables and means (and standard deviations), or medians (with lower and upper quartiles) for continuous variables will be presented.

There will be no tests of statistical significance nor confidence intervals for differences between randomised groups on any baseline variable.

PRINTED COPY MAY NOT BE THE MOST UP TO DATE VERSION

4.3 Reporting of Losses to Follow-up and Withdrawals of Consent

Number (with percentage) of participants lost to follow-up/withdrawn consent together with reasons, if available, will be reported by intervention arm. Any deaths (and their causes) will be reported separately.

4.4 Treatment Compliance

Control group: participants in the control group will be assumed to be compliant with their allocated treatment as they will receive standard usual care.

Intervention group: participants in the SC intervention group will be classed as being compliant with their allocated treatment if they received the package of the intervention they asked for at the time of registration (i.e. having received at least one CBT session for participants who requested CBT, having received the booklet, having access to the SC Website or having received all 3 packages).

Levels of intervention received: to be defined based on whether participants received only the booklet or the website or all 3 components of the SC Intervention in comparison to what they asked for once they were allocated to this group.

4.5 Other Variables

- Number and percentage of HVs that were awarded CBT accreditation out of total number of HVs randomised to the SC Intervention. Additionally, HV time until CBT accreditation will also be summarised by site. Other data related to the delivery of CBT will also be reported descriptively (overall).

The variables listed below will be reported descriptively by treatment group:

- Sleep of parents/carers as measured by the total hours of sleep (night time) and total hours of sleep (including day-time sleeping) during the last 2 weeks prior to the 8 and 16 week follow-up assessments along with the overall sleep quality rating reported at these study visits.

- Time from baseline assessment until 8 and 16 weeks follow-up assessment, measured in weeks.

- Babies' age at the time of the parents/carers' registration, measured in weeks.

PRINTED COPY MAY NOT BE THE MOST UP TO DATE VERSION

5 Protocol Deviations

All Protocol Deviations will be descriptively presented by treatment arm.

5.1 Major Deviations

- Ineligibility post-randomisation
- Non-compliance with the SC Intervention:
    - Not having received at least one CBT session for participants who requested CBT
    - Not having received the booklet and/or access to the SC Website for participants who requested one or both of these packages
    - Having missing data for the Parent/Carer Evaluation Questionnaire at 8 weeks follow-up (worst-case scenario)
- Missing a first EPDS assessment score altogether or first EPDS assessment score is outside the valid window assessment (+/-1 week)*

* Please refer to section 9 for more details on rules to be implemented for assessments carried out outside the Protocol timeframe.

5.2 Minor Deviations

- Week 8 visit/assessment attended outside of time window or a component missing (other than the primary outcome)
- Week 16 visit/assessment not performed as per Protocol:
    - Did not attend/unsuccessful contact
    - Attended/contacted out of time frame (i.e. time window or assessment deviations)
    - Component of assessment/procedure not done
- Participant not notified of allocation within the timeframe stated in the Protocol (1 working day)
- Time between first and last CBT session exceeds 6 weeks

The number and proportion of protocol deviations by type as well as the number and proportion of participants affected per deviation type will be reported by treatment arm and overall.

PRINTED COPY MAY NOT BE THE MOST UP TO DATE VERSION

6 Definition Of Population for Analyses

6.1 Complete-Case Population

The primary analysis will analyse complete cases (i.e. missing data will not be replaced) and participants will be included in the group to which they were randomised irrespective of the intervention received.

The Complete-Case population will consist of all participants recruited into the Trial who have a valid EPDS score at both baseline and 8 weeks follow-up. Any EPDS measurements taken up to (and including) 12 weeks post-baseline will be included in the analysis. The primary analysis of the primary outcome will follow the intention-to-treat principle, that is, participants will be analysed in the treatment group to which their HV team (cluster) was randomised regardless of whether participants were compliant with the Protocol and/or their allocated treatment.

6.2 Intention-to-Treat Population

The Intention-To-Treat (ITT) population will consist of all clusters randomised and will include all participants who were recruited into the Trial. Clusters will be attributed to the treatment arm to which they were randomised. Participants will be analysed in the treatment arm to which the cluster they belong to was randomised, regardless of what or how much intervention they received or any Protocol Deviations.

Missing data will be imputed using multilevel multiple imputation methods in order to account for the clustered nature of the data.

We will use multilevel multiple imputation to deal with missing data in the following types of variables: outcomes, covariates in the analysis model and any additional auxiliary variables in the imputation model. It is important to note that missing outcome data will not be imputed for clusters that dropped-out or for those that did not collect valid outcome data for at least one participant. Please refer to section 8.2.1 for the description of how multiple imputation will be done both at a cluster and individual level.

6.3 CACE Population

The population that will be used for the Complier Average Causal Effect (CACE) analysis will consist of all participants recruited into the Trial classed into two categories: compliant or non-compliant with their allocated treatment. All participants in the Control group will be deemed compliant with their allocated treatment as they will receive Usual Standard Care. Participants in the SC Intervention group will be classed as being compliant or non-compliant with their allocated treatment in line with the criteria provided in section 5.1.

6.4 Safety Population

The Safety population will consist of all participants recruited into the Trial who received at least one of the components of the SC package (i.e. access to the SC website, a copy of the SC booklet or CBT sessions). It is important to note that crossovers can only be one way (i.e. participants allocated to SC intervention receiving standard usual care).

PRINTED COPY MAY NOT BE THE MOST UP TO DATE VERSION

7 Outcome Measures

7.1 Primary Outcome

7.1.1 Definition and Derivation of Primary Outcome

Edinburgh Postnatal Depression Scale (EPDS) score is a measurement of parental depression measured at 8 weeks post baseline measures.

The EPDS score is defined as a continuous outcome that will be measured at 8 weeks post-baseline. The score is derived from a scale which comprises a list of 10 items, with each item being measured on a 4-point scale (0-3). EPDS scores range from 0-30 (inclusive) (Cox,1987).

7.1.2 Hypothesis to be investigated

The primary alternative hypothesis is that SC intervention is more efficacious in reducing the EPDS score at 8 weeks post-baseline compared to standard usual care.

7.2 Secondary Outcomes

7.2.1 Definition and Derivation of Secondary Outcomes

Edinburgh Postnatal Depression Scale (EPDS) score 16 weeks post baseline; proportion of parents/carers with EPDS score of 10 or more.

The EPDS score measured at 16 weeks post-baseline as a continuous measure in addition to the proportion of parents/carers with EPDS score of 10 or more at 16 weeks post baseline will be analysed as secondary outcomes. The continuous score measured at 16 weeks will be used to derive a binary variable that will represent the proportion of parents/carers with a score of 10 or more at this follow-up visit. Participants with an EPDS score greater than or equal to 10 will be classed as parent/carer may have minor/major depression whereas those with a score of less than 10 will be classed as parent/carer may not have minor/major depression.

Generalised Anxiety Disorder Assessment GAD-7 Scale self-rated measure of anxiety measured at 8 and 16 weeks post baseline.

The GAD-7 score is defined as a continuous outcome that will be measured at 8 weeks and 16 weeks post-baseline. The score is derived from a scale which comprises a list of 7 items with each item being measured on a four-point scale (0-3). GAD-7 scores range from 0 to 21 (inclusive). .

The levels of anxiety based on the GAD-7 cut off scores are defined as follows: “Minimal” for scores between 0-4, “Mild” for scores between 5-9, “Moderate” for scores between 10-14 and “Severe” for scores between 15-21 (Spitzer et al, 2006).

This secondary outcome will be derived as a categorical variable at 8 weeks and 16 weeks.

PRINTED COPY MAY NOT BE THE MOST UP TO DATE VERSION

EuroQol-5D-5L, self-rated health utility and quality of life measured at 8 and 16 weeks post baseline.

The EQ-5D-5L is defined as a continuous outcome that was designed to measure health-related quality of life across five domains: mobility, self-care, usual activities, pain/discomfort and anxiety/depression. Each domain or dimension has 5 levels of response: no difficulties, slight difficulties, moderate difficulties, severe difficulties and extreme difficulties scored on a scale from 1 to 5. Responses recorded for these domains using the EQ-5D-5L questionnaire will be used to map (or cross-walk) the 5L descriptive system data onto the 3L value set in accordance with new guidance published by NICE in January 2022 (Hernández-Alava et al, 2022).

Each of the five components of this secondary outcome will be derived as a categorical variable at 8 weeks and 16 weeks.

Parental Confidence Scale (MCQ) measures the confidence of a parent/carer in their parenting role at 8 and 16 weeks post baseline.

The MCQ is defined as a continuous outcome that will be measured at 8 and 16 weeks post-baseline. The score is derived from a scale which comprises a list of 14 items, with each item being measured on a 5-point scale (1-5). MCQ scores range from 14-70 (inclusive) (Parker,1985).

Each of the 14 items of this secondary outcome will be derived as a categorical variable at 8 and 16 weeks post-baseline.

Crying Knowledge Scale measures knowledge of infant crying, amount of perceived infant crying at 8 week post baseline follow-ups.

The Crying Knowledge Scale is defined as a categorical outcome that will be measured at 8 weeks post-baseline. This outcome comprises a list of 8 items with each item being measured on a 6-point scale (1-6) (Barr,2009).

Each of the 8 items of this secondary outcome will be derived as a categorical variable at 8 weeks post-baseline.

Parents/carers' ratings of how frustrating the crying is for them at 8 weeks follow-up.

The parents/carers' ratings of how frustrating the crying is for them is defined as a categorical outcome that will be measured at 8 weeks follow-up post-baseline. This outcome is measured on a 5-point scale (1-5) where the lowest and highest categories indicate "Extremely frustrated" and "Not frustrated at all" with the crying.

Parents/carers' ratings of how problematic the crying is for them at 8 weeks follow-up.

The parents/carers' ratings of how problematic the crying is for them is defined as a categorical outcome that will be measured at 8 weeks follow-up post-baseline. This outcome is measured on a 5-point scale (1-5) where the lowest and highest categories indicate "A severe problem" and "No problem at all" with the crying.

PRINTED COPY MAY NOT BE THE MOST UP TO DATE VERSION

Partner Relationship Quality (CSI-4) measures the relationship satisfaction at 8 and 16 weeks post baseline.

The Partner Relationship Quality is defined as a categorical outcome that will be measured at 8 and 16 weeks post-baseline. This outcome comprises a list of 5 items, with 4 of them being measured on a 6-point scale (1-6) and the 5th one being measured on a 7-point scale (1-7). Higher scores indicate higher levels of relationship satisfaction (Funk,2007).

Each of the 5 items of this secondary outcome will be derived as a categorical variable at 8 and 16 weeks post-baseline.

Health, breast–feeding maintenance will measure parental and infant health, breast-feeding persistence at 8 weeks and 16 weeks post baseline.

Measures of infant's crying, feeding and health will be assessed at 8 and 16 weeks post-baseline. Each of the measures that address this secondary outcome will be reported descriptively at 8 and 16 weeks post-baseline.

7.2.2 Hypotheses to be investigated

All secondary outcomes are hypothesised to benefit from the Intervention.

7.3 Exploratory Outcomes

There are no exploratory outcomes in line with the Protocol v9.0. Please see section 10.1 for details on exploratory analyses of the primary outcome.

7.4 Health Economic Outcomes

Health economic measures will be assessed at 8 weeks and 16 weeks post baseline. Contacts with the NHS; parents' use and parents' and HVs' ratings of services to be collected and measured.

Details regarding the definition and analysis of the health economic outcomes listed above are provided elsewhere in the HEAP.

PRINTED COPY MAY NOT BE THE MOST UP TO DATE VERSION

8 Primary Outcome Analysis

8.1 Primary Analysis of Primary Outcome

8.1.1 Primary Estimand

An estimand is a clear and unambiguous description of exactly what treatment effect is to be estimated in an RCT. It consists of five inter-related attributes: the population, the treatments (you want to compare), the outcome or endpoint, how to account for intercurrent events and a population-level summary measure of how the outcome between the different treatment conditions will be compared.

Estimand attribute	Description
Population	Parents who raise concern about their infants excessive crying up to age 20 weeks
Treatment(s)	Surviving Crying intervention is the offer of a leaflet, website and 4 optional CBT session with a specially trained HV alongside care as usual by routine HV services versus care as usual by routine HV services alone.
Outcome (endpoint)	EPDS score at 8 weeks post baseline/enrolment.
Handling Inter-Current events (ICEs)	1. Refusing or stopping SC intervention – treatment policy 2. SC arm participants switching to the usual care only arm – treatment policy 3. Not receiving SC intervention at all or only receiving parts of the SC intervention - treatment policy 4. Moving areas or HV teams – treatment policy 5. Use of other medications/treatments/therapy - treatment policy
Summary measures	Mean, median and skewness of the EPDS score in SC intervention compared to usual care. These summary measures will be calculated by estimating the shape (α) and scale (β) parameters of the EPDS score at 8 weeks post baseline/enrolment.

Strategies;
Treatment policy – regardless of any post randomisation events, the treatment effect is described from the final outcome measure in all patients with primary outcome data

Question answered by the estimand:
In parents who express concern over their infants' excessive crying and remain in trial follow-up and provide a valid EPDS at or around 8 weeks, what is the difference in mean, median and skewness of the EPDS score at 8 weeks (or at most before 12 weeks) between participants in clusters randomised to receive the offer of the SC intervention versus care as usual regardless of trial intervention's compliance/ discontinuation?

PRINTED COPY MAY NOT BE THE MOST UP TO DATE VERSION

8.1.2 Statistical Model

The primary analysis will be performed using generalised estimating equations (GEE), which require a minimum of 40 clusters in order to produce reliable estimates. An Analysis of Covariance (ANCOVA) approach will be used with each participant's continuous EPDS score at 8 weeks follow-up as the outcome variable, adjusting for their EPDS at baseline and for the baseline cluster mean, with GEE to allow for differences between clusters. The model will also include a categorical variable for randomisation group and terms for the stratification factors. The GEE model specification will include a log link function and a Gamma family distribution assumed for the outcome.

The primary analysis will analyse complete cases (i.e. missing data will not be replaced) and participants will be included in the group to which they were randomised irrespective of the intervention received.

The primary analysis model will also include HV teams (clusters) as a random effect to model participants' heterogeneity within sites. It will also include robust standard errors in order to provide consistent parameter and standard error estimates.

In addition to the log link function, the GEE model will also assume an exchangeable correlation structure (which specifies the within-group correlation structure) as well as the independence of clusters, a large sample size and that outcome data are missing completely at random (MCAR). An alternative link function, correlation structure and/or missing data mechanism will be considered if any of the assumptions are not satisfied.

The primary analysis of the primary outcome will use the Complete-Case population.

The model will only include participants with a valid EPDS score both at baseline and 8 weeks follow-up. Any participants whose EPDS measurements were taken up to (and including) 12 weeks post-baseline will be included in the model. This includes participants with no 8 week follow-up data collected and whose 16 week follow-up data was collected within the 12 week post-baseline window. If there is a numerical discrepancy in relation to time from baseline until 8 week follow-up assessment between the treatment groups, a variable of time until 8 week follow-up assessment will be included as well in the model.

The estimate of the multiplicative change in the average difference in EPDS score at 8 weeks follow-up between the intervention and control arm along with a two-sided 95% confidence interval and a p-value will be presented. Statistical significance will be considered at less than or equal to 5%. Furthermore, the intra-class correlation coefficient (ICC) will also be presented alongside a 95% confidence interval to assess the strength of the clustering effect. An ICC will also be estimated for a model with only the randomisation variable as a covariate as well as for a model without any covariates.

8.2 Secondary Analyses of Primary Outcome

Secondary analyses will repeat the analysis on both an intention-to treat and per-protocol basis.

PRINTED COPY MAY NOT BE THE MOST UP TO DATE VERSION

8.2.1 Intention to Treat Analysis

A secondary analysis of the primary outcome will be performed using the Intention-to-treat population. The ITT analysis will be conducted using the same model as the primary analysis of the primary outcome and will be adjusted for the same variables. Outcome data will be imputed where missing, collected out of timeframe or is not valid. Any EPDS measurement taken up to (and including) 12 weeks post-baseline will be included in the analysis model.

If there is a numerical discrepancy in relation to time from baseline until 8 week follow-up assessment between treatment groups, a variable of time until 8 week follow-up assessment will be included as well in the model.

The ITT analysis will be performed using multilevel multiple imputation (MMI) in R with the jomo package. The steps below will be implemented as part of the MI process:

- The imputation model will register imputation of the EPDS score variable at 8 weeks follow-up. The imputation model will also have auxiliary variables as outcomes if they have missing data or as covariates if they do not have any missing data.
The auxiliary variables are gender, ethnicity, participant's age, site, EPDS score at baseline.

- The MI will be multilevel with HV team ID as the cluster level variable.

- The multilevel MI in the jomo package in R is carried out using a joint modelling (JM) approach and with JM, for individuals for which one or more (but not all) outcome in the imputation model is missing, the imputation is carried out from the conditional distribution for one element of a multivariate normal model given the others. This means that in addition to the covariates specified, each of the outcome variables in the imputation model will also inform the imputation of the other outcomes if that information is available.

- The multilevel MI will use 20 imputations, 10,000 burn-in iterations and 10,000 between-imputation iterations and the imputations will be carried out separately by intervention arm [11]. A seed will be set in order to make the results reproducible.

- Once the imputations are carried out, the same model as the primary analysis of the primary outcome will be estimated using the lmer command in R. The model will be fitted for each of the 20 imputed datasets and then the estimates will be combined using Rubin's rules. For reproducibility, 95% confidence intervals on the basis of the final combined parameter estimates will subsequently be computed using the confint command.

The estimate of the multiplicative change in the average difference in EPDS score at 8 weeks follow-up between the intervention and control arm along with a two-sided 95% confidence interval and a p-value will be presented. Statistical significance will be considered at less than or equal to 5%. Furthermore, the intra-class correlation coefficient (ICC) will also be presented to assess the strength of the clustering effect. An ICC will also be estimated for a model with only the randomisation variable as a covariate as well as for a model without any covariates.

PRINTED COPY MAY NOT BE THE MOST UP TO DATE VERSION

8.2.2 CACE Analysis

A CACE analysis will be performed as a secondary analysis of the primary outcome instead of the Per-Protocol analysis. This analysis will be performed using the Complier Average Causal Effect (CACE) approach in order to estimate the treatment effect on participants who adhered to their allocated treatment (compliant) vs those who did not adhere (non-compliant).

For determining the treatment effect in the compliant population using the CACE analysis, compliance with the allocated treatment will be determined for each participant. As stated in section 6.1, participants allocated to the Control group will be deemed compliant. Participants in the SC Intervention group will be classed as being compliant with their allocated treatment if they received the package of the intervention they asked for at the time of registration (i.e. having received at least one CBT session for participants who requested CBT, having received the booklet, having access to the SC Website or having received all 3 packages). However, participants in this group will be classed as non-compliant if any of the conditions listed in section 5.1 are met.

The treatment effect estimation will be based on an instrumental variable (IV) approach using randomisation as the instrument. This method relies on standard assumptions: no defiers (monotonicity), randomisation affects outcomes only through the treatment received, and treatment assignment is independent of potential outcomes.

The CACE analysis will be conducted within a structural equation modelling (SEM) framework, implemented using Stata's gsem command, with random assignment serving as the instrumental variable for treatment received. The analysis model will be adjusted for the same variables used in the primary analysis. Any EPDS measurement taken between 7 and 9 weeks (inclusive) post-baseline will be included in the analysis model.

In this approach, the treatment adherence variable will be coded as 1 for compliants and 0 for all others, including non-compliant participants in the SC Intervention group.

The margins command will be used to report and interpret the results on the original scale. The estimate of the multiplicative change in the causal effect of treatment among compliant participants will be reported with a two-sided 95% confidence interval. Statistical significance will be considered at less than or equal to 5% level.

8.2.3 EPDS scores out of timeframe Analysis

An additional secondary analysis of the primary outcome that explores the impact of EPDS scores measured outside the Protocol timeframe on the SC Intervention will be conducted using the ITT population. If more than 5% of cases fall outside of window, observations will be limited to +/- 1 week window, with missing values being imputed.

This secondary analysis will be performed using multilevel multiple imputation (MMI) in R with the jomo package. The steps below will be implemented as part of the MI process:

- The imputation model will register imputation of the EPDS score variable at 8 weeks follow-up. The imputation model will also have auxiliary variables as outcomes if they have missing data or as covariates if they do not have any missing data. The auxiliary variables are gender, ethnicity, participant's age, site, EPDS score at baseline.

PRINTED COPY MAY NOT BE THE MOST UP TO DATE VERSION

- The MI will be multilevel with HV team ID as the cluster level variable.

- The multilevel MI in the jomo package in R is carried out using a joint modelling (JM) approach and with JM, for individuals for which one or more (but not all) outcome in the imputation model is missing, the imputation is carried out from the conditional distribution for one element of a multivariate normal model given the others. This means that in addition to the covariates specified, each of the outcome variables in the imputation model will also inform the imputation of the other outcomes if that information is available.

- The multilevel MI will use 20 imputations, 10,000 burn-in iterations and 10,000 between-imputation iterations and the imputations will be carried out separately by intervention arm [11]. A seed will be set in order to make the results reproducible.

- Once the imputations are carried out, the same model as the primary analysis of the primary outcome will be estimated using the lmer command in R. The model will be fitted for each of the 20 imputed datasets and then the estimates will be combined using Rubin's rules. For reproducibility, 95% confidence intervals on the basis of the final combined parameter estimates will subsequently be computed using the confint command.

The estimate of the multiplicative change in the average difference in EPDS score at 8 weeks follow-up between the intervention and control arm along with a two-sided 95% confidence interval and a p-value will be presented. Statistical significance will be considered at less than or equal to 5%. Furthermore, the intra-class correlation coefficient (ICC) will also be presented to assess the strength of the clustering effect. An ICC will also be estimated for a model with only the randomisation variable as a covariate as well as for a model without any covariates.

8.3 Adjustment of P values for Multiple Testing

There will be no adjustment for multiple testing for the secondary, sensitivity, exploratory and subgroup analyses of the primary outcome or the secondary outcomes.

8.4 Missing Data

Missing data will be imputed using multilevel multiple imputation methods in order to account for the clustered nature of the data.

It is important to note that the multilevel multiple imputation method for handling missing data will be implemented only in the ITT analysis of the primary outcome. Please refer to section 8.2.1 for details of the implementation.

8.5 Pre-specified Subgroup Analysis

Subgroup analyses of the primary outcome will be performed on the following factors: SC HV accredited in CBT, social deprivation index and ethnicity.

Subgroup analyses of the primary outcome will be conducted on the social deprivation index, ethnicity and level of education. An exploratory analysis of the primary outcome will be conducted on the SC HV accredited in CBT instead of a subgroup analysis as it will not be a comparison between treatment groups. Details on this analysis can be found in section 10.1.

PRINTED COPY MAY NOT BE THE MOST UP TO DATE VERSION

Methodology similar to that proposed for the primary analysis of the primary outcome in section 8.1 will be implemented to assess the treatment effect across different subgroups of participants as outlined in section 1.2.4. This analysis will be conducted using the Complete-Case population.

To investigate the treatment effect in subgroups (social deprivation, ethnicity and level of education), we will utilise the same model used for the primary analysis of the primary outcome, considering each one of the factors mentioned above. For each subgroup analysed, the model, in addition to the variables included in the primary analysis, will include a categorical variable denoting which subgroup the individual belongs to and an interaction term between the treatment group and subgroup to assess the level of heterogeneity in treatment effect between the subgroups.

Separate models will be fitted to estimate the treatment effect in each subgroup and estimates and 95% confidence intervals presented in Forest plots (alongside tests of heterogeneity using the subgroup by treatment interaction term in the previously described model). Statistical significance will be considered at less than or equal to 5%.

8.6 Sensitivity Analysis

A sensitivity analysis of the primary outcome will adjust for ethnicity and social deprivation index at the individual level.

No longer applicable as this will be addressed by the subgroup group analysis that will be conducted on these factors.

9 Secondary Outcome Analyses

9.1 Primary Analysis of Secondary Outcomes

Secondary outcomes will be analysed using a complete case population, in a similar manner to the primary outcome, modelled assuming an appropriate distribution family and link function.

Continuous outcomes

1. Edinburgh Postnatal Depression Scale (EPDS) score 16 weeks post baseline

The primary analysis of this secondary outcome will be performed using generalised estimating equations (GEE). An Analysis of Covariance (ANCOVA) approach will be used with each participant's continuous EPDS score at 16 weeks follow-up as the outcome variable, adjusting for their EPDS at baseline and for the baseline cluster mean, with GEE to allow for differences between clusters. The model will also include a categorical variable for randomisation group and terms for the stratification factors. The GEE model specification will include a log link function and a Gamma family distribution assumed for the outcome.

The primary analysis will use the Complete-Case population (i.e. missing data will not be replaced) and participants will be included in the group to which they were randomised irrespective of the intervention received.

PRINTED COPY MAY NOT BE THE MOST UP TO DATE VERSION

The primary analysis model will also include HV teams (clusters) as a random effect to model participants' heterogeneity within sites. It will also include robust standard errors in order to provide consistent parameter and standard error estimates.

In addition to the log link function, the GEE model will also assume an exchangeable correlation structure (which specifies the within-group correlation structure) as well as the independence of clusters, a large sample size and that outcome data are missing completely at random (MCAR). An alternative link function, correlation structure and/or missing data mechanism will be considered if any of the assumptions are not satisfied.

The model will only include participants with a valid EPDS score both at baseline and 16 weeks follow-up. Any participants whose EPDS measurements were taken after 12 weeks post-baseline will be included in the model. If there is a numerical discrepancy in relation to time from baseline until 16 week follow-up assessment between the treatment groups, a variable of time until 16 week follow-up assessment will be included as well in the model.

The estimate of the multiplicative change in the average difference in EPDS score at 16 weeks follow-up between the intervention and control arm along with a two-sided 95% confidence interval and a p-value will be presented. Statistical significance will be considered at less than or equal to 5%. Furthermore, the intra-class correlation coefficient (ICC) will also be presented alongside a 95% confidence interval to assess the strength of the clustering effect. An ICC will also be estimated for a model with only the randomisation variable as a covariate as well as for a model without any covariates.

2. Generalised Anxiety Disorder Assessment GAD-7 Scale self-rated measure of anxiety measured at 8 and 16 weeks post baseline

3. EuroQol-5D-5L, self-rated health utility and quality of life measured at 8 and 16 weeks post baseline

4. Parental Confidence Scale (MCQ) measures the confidence of a parent/carer in their parenting role at 8 and 16 weeks post baseline

The secondary outcomes listed above will be analysed by implementing a parametric approach using the Complete-Case population. Prior to the implementation of this approach, an assessment of normality will be carried out to investigate whether the assumptions are met. In case these are not satisfied, an alternative distribution family will be considered and implemented.

The outcomes will be compared between treatment groups using Mixed Effects Linear Regression models, with each model adjusting for their respective baseline measure. Treatment group and site (stratification factor) will be included as fixed effects in the models while the HV teams (clusters) will be included as a random effect to model participants' heterogeneity within sites. These models will only include participants with valid outcome data at both baseline and follow-up.

For untransformed outcome data, treatment comparison estimates will be presented as adjusted means with two-sided 95% confidence intervals. However, for transformed data, treatment comparisons will be presented as adjusted geometric means with two-sided 95% confidence intervals. Statistical significance is considered at less than or equal to 5%.

PRINTED COPY MAY NOT BE THE MOST UP TO DATE VERSION

The residuals of the model that will be fitted for each of the outcomes will be assessed for normality by performing the Shapiro-Wilk test. In the event that the assumptions of the Mixed Effects Linear Regression are not met, another family distribution will be considered or, alternatively, a non-parametric model will be fitted instead.

Binary outcomes

1. Proportion of parents/carers with Edinburgh Postnatal Depression Scale (EPDS) score of 10 or more at 16 weeks post-baseline

The proportion of parents/carers with a valid EPDS score of 10 or more at 16 weeks will be analysed by implementing a parametric approach using the Complete-Case population. This binary outcome will be compared between treatment groups using a Multilevel Binary Logistic Regression model. Treatment group and site (stratification factor) will be included as fixed effects in the analysis model while the HV teams (clusters) will be included as a random effect to model participants' heterogeneity within sites.

Treatment comparison estimates will be presented as adjusted Odds Ratios (OR) and two-sided 95% confidence intervals. Statistical significance is considered at less than or equal to 5%.

In the event that the assumptions of the Multilevel Logistic Regression are not met, a non-parametric approach will be implemented instead.

Categorical outcomes

1. Crying Knowledge Scale at 8 weeks post-baseline

2. Parents/carers' ratings of how frustrating the crying is for them at 8 weeks post-baseline

3. Parents/carers' ratings of how problematic the crying is for them at 8 weeks post-baseline

4. Partner Relationship Quality (CSI-4) at 8 and 16 weeks post-baseline

5. Health, breast–feeding maintenance at 8 weeks and 16 weeks post-baseline

The secondary outcomes listed above will be analysed by implementing a parametric approach using the Complete-Case population. The outcomes will be compared between treatment groups using a Multilevel Binary, Multinomial or Ordinal Logistic Regression model. Treatment group and site (stratification factor) will be included as fixed effects in the analysis models while the HV teams (clusters) will be included as a random effect to model participants' heterogeneity within sites.

The outcome of breastfeeding maintenance will be analysed on a subset consisting of participants who reported breastfeeding their infant at Baseline. A binary outcome will be defined to classify participants who still breastfed their infant at 8 and 16 weeks follow-up post-baseline. A summary of how many participants breastfed from the beginning will be presented by treatment group.

PRINTED COPY MAY NOT BE THE MOST UP TO DATE VERSION

Treatment comparison estimates will be presented as adjusted Odds Ratios (OR) and two-sided 95% confidence intervals. Statistical significance is considered at less than or equal to 5%.

In the event that the assumptions of the Multilevel Logistic Regression are not met, a non-parametric approach will be implemented instead.

9.2 Secondary Analyses of Secondary Outcomes

Not applicable.

9.3 Missing Data

As the complete case population will be used in the primary analysis of the secondary outcomes, missing data will not be replaced and participants will be included in the group to which they were randomised irrespective of the intervention received.

9.4 Pre-specified Subgroup Analysis

Not applicable.

10 Additional Analyses

10.1 Exploratory Analyses of the Primary Outcome

Exploratory analyses of the primary outcome will investigate the effect of differing levels of compliance over and above the minimum level.

Exploratory analyses of the primary outcome to investigate the effect of differing levels of compliance will not be conducted as one level of intervention received is deemed sufficient as advised by the TMG. Summary statistics of the extent to which participants received different levels of intervention will be produced.

We will describe the EPDS score at 8 weeks follow-up in the Intervention group participants who received at least one CBT session split by whether the session was given by an accredited or non-accredited HV. This descriptive summary will be repeated for both definitions of accredited/unaccredited (i.e. based on at time of delivery of CBT sessions and based on at some point during the trial) outlined in section 2.1.

PRINTED COPY MAY NOT BE THE MOST UP TO DATE VERSION

11 Safety and Adverse Events Reporting

SC is not a clinical trial of an investigational medicinal product, therefore the usual monitoring of pharmacovigilance and associated terminology is not relevant.

Any Serious adverse events (any adverse event or adverse reaction that results in death, is life-threatening, requires hospitalisation or prolongation of existing hospitalisation, results in persistent or significant disability or incapacity, or is a congenital anomaly or birth defect), must be collected by the research team and reported to the sponsor within 24 hours of the research team being made aware of the event.
Throughout the study, participants will be asked about their infant's health at each follow up visit and be asked to report any change in health. Local safe-guarding procedures will be followed by the researcher if deemed appropriate by the Principal Investigator.

On identification of a potential serious adverse event (SAE), the relevant local Principal Investigator will be informed and will review the SAE to determine the likelihood that this is related to the study intervention or to the infant's excessive crying. It will then be reported through the study reporting system.

All safety data will be presented according to the Safety population.

All serious adverse events will be reported by treatment arm and overall based on the following characteristics: seriousness, duration (measured in days), relatedness, severity, action taken, outcome and expectedness.

PRINTED COPY MAY NOT BE THE MOST UP TO DATE VERSION

12 References

Cox, J.L., Holden, J.M., Sagovsky, R. (1987) Detection of postnatal depression: Development of the 10-item Edinburgh Postnatal Depression Scale. British Journal of Psychiatry, 150, 782-786

Spitzer, R.L., Kroenke, K., Williams, J.B.W., Löwe, B. A. (2006). Brief Measure for Assessing Generalized Anxiety Disorder: The GAD-7. *Arch Intern Med.* 166(10):1092–1097. doi:10.1001/archinte.166.10.1092

Hernández-Alava, M., & Pudney, S. (2022). Mapping between EQ-5D-3L and EQ-5D-5L: A survey experiment on the validity of multi-instrument data. *Health Economics*, 1–17. https://doi.org/10.1002/hec.4487

Parker S. & Zahr L.K., (1985). The maternal confidence questionnaire. Boston: Boston City Hospital

Barr RG, Barr M, Fujiwara T, et al. Do educational material change knowledge and behaviour about crying and shaken baby syndrome? A randomised controlled trial. CMAJ 2009;180. DOI:10.1503/cmaj.081419. Copyright2009 Canadian Medical Association

Funk,J.L., & Rogge,R.D..(2007). Testing the the ruler with Iten response Theory: Increasing precision of Measurment for Relationship Satisfaction with Couples Satisfaction Index. Journal of Family Psychology,21,572-583

PRINTED COPY MAY NOT BE THE MOST UP TO DATE VERSION

13 Appendices

13.1 Appendix 1: Templates for Tables, Listings and Figures

PRINTED COPY MAY NOT BE THE MOST UP TO DATE VERSION