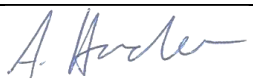
An individual-level randomised controlled trial of group antenatal care

Statistical Analysis Plan

Version: 1.0
Date: 02/JUL/2024

Person(s) contributing to the analysis plan	
Name(s) and position(s)	Olivier Quintin Thomas Hamborg Angela Harden Bethan Hatherall Liz Allen
Authorisation	
Position	**Chief or principal investigator**
Name	Professor Angela Harden
Signature	
Date	01/07/2024
Position	**Senior trial statistician**
Name	Thomas Hamborg
Signature	Thomas Hamborg Digitally signed by Thomas Hamborg Date: 2024.07.02 00:31:57 +01'00'
Date	
Position	**Independent statistician**
Name	Liz Allen
Tick once reviewed	✓
Date	**15/12/2023**

1. Administrative Information

1.1 Trial registration number: 91977441

This SAP is based on protocol version 10.0 (date 06/06/2024)

1.2 SAP revision history

Protocol version	Updated SAP version no.	Section number changed	List of changes from previous version/protocol	Author of change	Date
4.0	0.1		Initial draft started	CM	01/07/2020
4.0	0.2	all	commented and added on all sections of the SAP	TH	20/08/2020
4.0	0.3	all	Continued progress on SAP; focus on analysis section 5	CM	21/08/2020
4.0	0.4	Analysis / outcomes	Focussed again on section 5	CM	28/08/2020
4.0	0.5	analysis/outcomes	Focus on updating outcomes, and adding information based on study group feedback for missing data/follow-up	CM	17/09/2020
4.0	0.6	All	Added section 5.10, 6.1, 6.2. Amended various other sections	TH	22/09/2020
6.0	0.7	All	Added to sections on routine maternity data, updated data cleaning and sensitivity analyses section and amended various other sections	CM	28/09/2020
6.0	0.8	All	Began addressing TSC independent statistician (LA) comments about SAP	CM	08/12/20

			and continuing to update		
6.0	0.9	All	Filled in some additional information based on further feedback from study team. Added notes about multiple pregnancy analysis and looked further into compliance	CM	12/05/2021
8.0	0.91	all	Comments on primary outcome analysis and estimands framework.	OQ	03/2022
8.0	0.92	all	Incorporate the learning from reviewing the data to produce the data snapshot; refine the definition of the primary and SVB outcome measures; describe a strategy for using data affected by the pandemic and create template tables.	OQ	23/09/2022
8.0	0.93		Change in the definition of the primary outcome; change in the study population; consideration added for secondary outcomes.	OQ	24/07/2023
8.0	0.94		Change after Thomas's review.	OQ	21/09/2023
9.0	0.10		Change after Liz's review.	OQ	11/01/2024
9.0	0.11		Development of the SAP following meetings	OQ	13/05/2024

9.0	0.12		Update after comments from trial team	OQ	03/06/2024
10.0	0.13		Update after comments from senior statistician	OQ/TH	01/07/2024
10.0	1.0		Sign off		01/07/2024

*If the SAP has been published, indicate which version.

1.3 Members of the writing committee

Sandra Eldridge was responsible for the original statistical analysis strategy in the protocol. Connor Mustard and Olivier Quintin have written the statistical analysis plan under the direction of Tom Hamborg. Angela Harden, Meg Wiggins, Mary Sawtell and Lorna Sweeney have also contributed to the writing of this statistical analysis plan.

1.4 Timing of SAP revisions in relation to unblinding of data/results

This document has been developed prior to examination of unblinded trial data by those contributing.

1.5 Analysis software

All analyses and data presentation described in this document and will be carried out using Stata version 18.0 or later unless otherwise specified.

1.6 Remit of SAP

REACH is a programme grant. This SAP covers outcomes from work package 3 (Pregnancy Circles Trial). This plan is intended not to change or contradict the general aims of the protocol, but rather expand on them. In the event of a discrepancy the analyses described here will supersede those in earlier documents.

2. Background and trial design

Study objectives	*This trial study objectives are as follows:* *a. To assess whether Pregnancy Circles (group-based antenatal care) improves the health of babies compared with the standard individual model of antenatal care.* *b. To assess whether attending Pregnancy Circles improves maternal outcomes such as empowerment and post-natal depression, as well as increasing women's satisfaction with antenatal care.* *c. To assess cost-effectiveness, intervention mechanisms, and acceptability of group-based antenatal care to women and staff and issues relevant to future sustainability and wider implementation in the NHS.*
Study design	*Individually randomised, parallel group, randomised controlled superiority trial with integrated process and economic evaluations*
Setting	*Multi-centre study across 14 NHS Trusts, namely Barts Health NHS Trust, Whittington Hospital NHS Trust, Lewisham & Greenwich NHS Trust, West Hertfordshire NHS Trust, East Suffolk and North Essex NHS Foundation Trust, Princess Alexandra Hospital NHS Trust, Worcestershire Acute Hospitals NHS Trust,* Mid and South Essex NHS Foundation Trust*, Ashford and St Peter's Hospitals NHS Foundation, East Sussex Healthcare NHS Trust, Lancashire Teaching Hospitals NHS Foundation Trust, Surrey and Sussex Healthcare NHS Trust, Epsom & St. Helier University Hospitals NHS Trust and Royal Free London NHS Foundation Trust*
Participants	*Women who are currently pregnant and registered for antenatal care with the included NHS Trust maternity services, whose estimated due dates fit with the proposed group start dates, and who live within the usual working areas of these services*
Interventions	*Intervention* *Pregnancy circles consisting of 8-12 pregnant women, facilitated by two midwives (and supplemented with interpreters and/or other support staff as appropriate). There will be a total of eight antenatal group sessions each of which will last for approximately two hours. The first part of each session will involve "self-care activities" (ex. Women will be encouraged to take an active part in their antenatal care by testing their own urine, taking their own/each other's blood pressure and writing the results in their notes). Following these checks, the sessions will involve short one-to-one sessions with one of the midwife facilitators for individual health checks (ex. Abdominal pain) which will take place on a mat in the corner of the room while the rest of the group has group discussion facilitated by the second midwife. Women will be allowed to request more privacy for one-to-one time. The women in the group will also be invited to one post-natal reunion session.* *Control* *Usual antenatal care in the maternity service*

Primary outcome measure(s)	*A "Healthy Baby" composite outcome consisting of the 4 following components:* 1. *Live baby (i.e. no stillbirth after 24 completed weeks of pregnancy, no miscarriage before 24 completed weeks and no neonatal death within 28 days of the birth)* 2. *Born at term (≥ 37 weeks + 0 days)* 3. *Appropriate weight for gestational age (GROW centile>9.99 & <90.01)* 4. *Not admitted to a neonatal care unit (which includes: Intensive Care Unit, SCBU and High Dependency Unit. But NOT transitional care)*

3. Outcome Measures

3.1 Timing of outcome measures

Data on the various outcome measures used in this study will be collected either at baseline (recruitment; during the first antenatal booking appointment/first dating-scan appointment), first follow-up (35-weeks gestation), birth (routine maternity data) or second follow-up (3-months postnatal) or a combination of these time points. The table below shows the outcomes, the validated scales/measures which are being used where appropriate, and the specific time points when outcome data will be collected.

Outcome measure and data collection timing

Outcome	Validated measure (where applicable)	Baseline	First follow up - 35 weeks gestation	Birth – routine maternity data	Second follow up - 3 months postnatal
Live baby				✓	
Born at term				✓	
Appropriate weight for gestational age	GROW centile >9.99 & <90.01			✓	
Not admitted to a Neonatal Care Unit				✓	
Spontaneous vaginal delivery (SVD)				✓	✓
Women's empowerment	Pregnancy-related Empowerment Scale (PRES)		✓		
Women's satisfaction with maternity care	Friends and family test		✓		✓
Attendance at antenatal care			✓	✓	
Social support	The Duke-UNC Functional Social Support Questionnaire	✓			✓
Self-efficacy	Pearlin Mastery Scale	✓	✓		✓

Prenatal stress	Revised prenatal distress scale	✓	✓		
Caesarean delivery				✓	✓
Infant birth weight (g)				✓	
Place of birth				✓	
Breast feeding initiation					✓
Breast feeding continuation and exclusivity					✓
Postnatal depression	Edinburgh Postnatal Depression Scale (EPDS)				✓
Health Literacy	Health Literacy Questionnaire (HLQ) (1 domain)		✓		
Postnatal symptoms	**(***NPEU checklist***)**				✓
Emotional wellbeing	*Short Warwick-Edinburgh Mental Wellbeing Scale (SWEMWBS)*	✓	✓		✓
Health related quality of life	EQ5D-5L	✓	✓		✓
Continuity with care			✓		
Choice in care			✓		✓
Involvement in care					✓
Preparedness for labour and birth			✓		✓
Confidence in caring for baby after birth					✓
Immunisation					✓

Additional measures of satisfaction with care			✓		✓

3.2 Primary outcome

The "healthy baby" composite consists of the following 4 components:

1. Live baby (i.e. no pregnancy loss before 24 completed weeks, no stillbirth after 24 completed weeks of pregnancy and no neonatal death within 28 days of the birth)
2. Born at term (≥ 37 weeks + 0 days)
3. Appropriate weight for gestational age (GROW centile >9.99 & <90.01)*
4. Not admitted to a neonatal care unit (which includes Intensive Care Unit, SCBU and High Dependency Unit, but NOT transitional care)

A baby is considered a "healthy baby" only if the answer to all above questions is "yes", otherwise "no" (binary outcome measure). The primary outcome will be considered missing if any of its components are missing apart from the following exception: If Live baby is recorded as 'no' then the healthy baby outcome is 'no' regardless of whether other components are missing.

** The GROW centile macro version 8.0.6.2 will be used for component 3. Appropriate weight for gestational age. It is multidimensional and includes: maternal height and weight, ethnic origin, parity at booking, gestational age, baby alive (yes/no), baby gender and weight (g).*

Note: at the grant application stage, the secondary outcome, spontaneous vaginal delivery (SVD) was defined as the primary outcome for the trial. However, following advice from and discussion with various stakeholders the "healthy baby" composite outcome was deemed a more suitable primary outcome for this project and chosen prior to commencement of the trial.

3.3 Secondary Outcomes*:

1. Spontaneous vaginal delivery (SVD)** defined as a woman who delivers vaginally (binary: for SVD, yes favoured)
2. Women's empowerment using pregnancy-related empowerment scale (continuous: sum of individual items with scores ranging from 16-64; higher scores are favoured)
3. Women's satisfaction with maternity care using *NHS* Friends and family test (continuous: score on one question with scores ranging from 1-5; lower scores favoured)
4. Breast feeding initiation (binary: did mother ever initiate breastfeeding, yes favoured --> however, possible different responses for descriptive analysis)
5. Mental wellbeing using Short Warwick-Edinburgh Mental Wellbeing Scale (continuous: total score ranges from 7-35, higher scores indicate higher positive mental wellbeing)
6. Live baby (i.e. no pregnancy loss before 24 completed weeks, no stillbirth after 24 completed weeks of pregnancy and no neonatal death within 28 days of the birth) using health records (binary: yes favoured)
7. Born at term (37 weeks + 0 days and above) using health records (binary: yes favoured)

8. Appropriate weight for gestational age (GROW centile >9.99 & <90.01) using health records (binary: yes favoured)
9. Not admitted to a neonatal care unit (which includes Intensive Care Unit, SCBU and High Dependency Unit, but NOT transitional care using health records binary: yes favoured)

** Those for whom the primary outcome composite healthy baby variable is 'no' will inevitably have a lower response rate for the first follow up questionnaire. This is because the composite includes whether a baby is alive and born at term (i.e. at 37+ weeks gestation). Participants who have experienced pregnancy loss (baby alive = no) are not sent the first (or second) follow up questionnaire. Participants who deliver prematurely (<35 weeks) will not have had the opportunity to complete the first follow up questionnaire before delivery and are unlikely to complete it after. Analyses of variables collected via the follow up questionnaire, such as empowerment, satisfaction with care, self-efficacy, pre-natal stress, health literacy and emotional well-being, will acknowledge the exclusion of these groups, particularly if it is found that the intervention does have a significant effect on the primary outcome.*

*** Routine data will be used as the default source of data for SVD. We will use second follow-up SVD data only if routine data values for a participant are missing.*

3.4 Additional Outcomes

Additional outcomes have been included due to a) interest from the study team and b) to reflect those intermediate outcomes in the Pregnancy Circles logic model which are not included as secondary outcomes (see Wiggins et al., 2020). Additional outcomes have been listed separately from secondary outcomes to avoid multiplicity issues due to an excessive number of secondary outcomes and are to serve predominantly for hypothesis generation.

1. Social support using Duke Social Support Scale (continuous: scores range from 8-40, higher scores favoured)
2. Self-efficacy using Perlin Mastery Scale (continuous: scores range from 7-28 with higher scores favoured)
3. Prenatal stress using Revised Prenatal Distress Questionnaire (continuous: 12-item total score, scores range from 0-16; higher is favoured)
4. Health literacy using Health Literacy Questionnaire (continuous: score on only first domain of HLQ, range 1-20, higher score favoured)
5. Attendance at antenatal care (continuous: number of sessions, higher is favoured)
6. Additional measures of satisfaction with care (categorical: (FU1) - Overall, how do you feel about the care you received from midwives? Options - very happy, fairly happy, not very happy, very unhappy; (FU2) Overall, how do you feel about the care you received from midwives (before the birth of your baby)? – options – very happy, fairly happy, not very happy, very unhappy)
7. Caesarean delivery (categorical: planned, emergency, none)
8. Infant birth weight in grams (continuous)
9. Place of birth (categorical: actual place of delivery; options= hospital obstetric unit, hospital alongside midwifery unit, freestanding midwifery unit, home or other)

10. Breast feeding continuation and exclusivity (binary: did mother exclusively breastfeed to 3-month follow-up, yes favoured --> however will look at exclusive, breast milk, artificial exclusive, mixed, other in descriptive analysis)
11. Postnatal depression using Edinburgh Postnatal Depression Scale (binary: mothers scoring above 13 are "likely to be suffering from a depressive illness of varying severity", no favoured)
12. Postnatal symptoms using NPEU (National Perinatal Epidemiology Unit) checklist (based on the NPEU checklist items to produce a single score variable; continuous; lower scores favoured from 0 to 5). The items include psychological symptoms (e.g. "the blues", depression, anxiety); posttraumatic stress-type symptoms (e.g. flash-backs, difficulties concentrating, sleep problems not related to the baby); bodily changes (e.g. stress incontinence, backache; difficulties/pain during intercourse); birth-related symptoms (e.g. painful stitches, wound infection); breastfeeding problems and severe fatigue.
13. Immunisation (categorical: Has your new baby had their routine immunisations at 2 months and 3 months of age? Options – yes 2 months; no 2 months; yes 3 months; no 3 months)
14. Continuity of antenatal care (categorical: how many midwives did you have during care; options=1-2, 3, 4+ or don't know with 1-2 being favoured) and satisfaction with continuity (categorical: Do you feel in general that the midwives you saw during your regular antenatal appointments got to know you and remembered you and your progress? Options – yes definitely; yes, a little; no, not really; no, not at all; don't know can't remember; How satisfied are you with how much midwives got to know you and remember you and your progress? Options – very satisfied; quite satisfied; not at all satisfied; don't know/can't remember; Do you feel that midwives have been sensitive to your cultural and/or language needs? Options – yes, definitely; yes, a little; no, not at all; don't know/can't remember)
15. Choice in care (categorical: Were you offered any of the following choices about where to have your baby? Options – a choice of different hospitals; in a midwife-led unit or a birth centre; in a consultant-led unit; at home; I was not offered any choices; I was not offered any choices due to medical reasons; don't know)
16. Involvement in care (categorical: Thinking about your regular antenatal care, do you think you were involved enough in decisions about your care? Options – yes, always; yes – sometimes; no and I wanted to be; no and I did not want to be; don't know)
17. Preparedness for labour and birth (categorical: How prepared did you feel for labour and birth? options: very well, quite well, not very well and not at all well and How well did you manage during labour? – options - very well, quite well, not very well and not at all well)
18. Confidence in caring for baby after birth (categorical: How confident did you feel about caring for your baby in the first week after the birth?: options – very confident, fairly confident, not very confident, not at all confident, don't know/can't remember and Have you received enough help and advice from a midwife and/or health visitor about your baby's health, care and progress? Options – yes, definitely; yes to some extent; no, and I wanted help/advice; no, but I did not need any; don't know)

4. Study methods

4.1 Sample size

4.1.1 Sample size calculation (pre covid)

For the primary outcome ("Healthy baby"), in order to detect a difference in babies born "healthy" of 8% between the control and intervention arm, with 90% power and a 5% significance level, we would require at least 866 women per arm (i.e. 1732 total). This calculation also assumed an outcome proportion of 69% in the control arm, accounts for clustering within the intervention arm with an intra-cluster correlation (ICC) of 0.05 (in the intervention arm), mean group sizes of 8 with cluster size variability assuming Poisson distribution and assumes 10% drop-out in both arms. This sample size provides 84.8% power to detect a difference between arms of 7.3% in our, former primary outcome, spontaneous vaginal birth. Thus, a sample size 1732 is sufficiently powered to detect changes in both the primary outcome and SVD (now a secondary outcome).

4.1.2 Sample size calculation (recruitment unpaused)

With the approval of the TSC an extension to allow the recruitment of another n=566 women to the study in addition to the sample size above has been made after the trial was paused. This results in a final total sample size targe of n=2190. The reasons for proposing this sample size are described in the following.

The trial has been paused to recruitment because of the Covid-19 pandemic. In person pregnancy circles were stopped and women were returned to one-to-one care. At this point:

- 794 women were recruited prior to the pandemic.
- 532 women recruited prior to the onset of the pandemic had not quite completed the intervention period but had the chance to receive a high or moderate dose of the intervention.
- 176 women recruited prior to the onset of the pandemic only had the opportunity to receive a low/very low dose of the intervention.
- Finally, another 122 women were recruited and had their pregnancy during a period when no intervention delivery was possible.

At this moment, the recruitment of 106 additional women would have been needed to reach the pre-specified sample size of 1732. But since some women only received low/very low or no dose of the intervention, we revised the sample size calculation to address the reduced chance of observing the pre-hypothesised effect size in the cohort already recruited.

With the approval of the TSC the decision was made to exclude all women who during the pandemic were not able to receive the intervention or were only able to receive a low/ very low dose during the pandemic from the primary outcome analysis (NB data from these women will still be analysed, just not as part of the primary outcome analysis). Additional women are recruited to replace these n=298 women. For the participant cohort who could have received a moderate or high dose of the intervention during the pandemic the assumption was made the treatment effect was halved which

requires an inflation of the sample size by n=160 to maintain pre-specified power. Thus together with 108 participants who had yet to be recruited the total post-pandemic recruitment target is n=108+160+298=566.

4.2 Randomisation procedure

Women will be randomised to receive current standard mid-wife led antenatal care (control) or entered in the pregnancy circles groups (intervention) with a 1:1 allocation ratio. If recruited in person, each woman will be told her allocation status face-to-face straight away. Each pregnancy circle will consist of 8-12 pregnant women who have estimated delivery dates within the same approximate one-month period. Women will be randomised using randomly permuted blocks of 4, 6 or 8 and randomisation will be stratified by the location (site) of the Pregnancy Circle and on a woman's ability to speak English a) well/very well or b) not well/not at all.

4.3 Blinding

Participants and maternity staff will be unblinded to allocations, along with the researchers conducting the process evaluation observations and interviews. However, data informatics staff supplying outcome information from electronic records and researchers accessing paper records for outcome information will be blinded to allocation. Furthermore, all individuals contributing to the analysis plan and those conducting the analysis at the PCTU will be blinded to intervention allocation until the formal SAP has been signed off and the database locked for analysis.

5. Analysis methods

5.1 Data cleaning process

Blinded data cleaning has been an ongoing process for the routine maternity data which includes the primary outcome amongst other data related to birth. These data were sent in excel files by sites and data cleaning included range, logical and consistency checks and identifying unreported Serious Adverse Events as defined in the protocol. The process of creating queries has been handled with Stata. Data cleaning reports were then sent to the trial manager to communicate with sites for data correction to ensure the quality of the data before it is locked for the final analysis. A summary of issues identified prior to sign off of the SAP is provided in Appendix Table 17

Data collected on REDCap are self-reported questionnaires that patients have completed throughout the trial. As it was not possible to get back to patients to eventually amend their responses, limited data cleaning is possible on these data. Consistency checks described below will be made and summarised in a table in an appendix to the statistical analysis report.

1) Consistency of English level between Randomisation data and baseline questionnaire.

2) Check that all randomised participants who have not withdrawn before randomisation have correctly been assigned to a pregnancy circle.

3) Consistency check: estimated delivery date should be equal to confirmed delivery date more or less 4 months.

4) If AE/SAE has a start date then there should be a description.

5) End date of AE/SAE must be equal or greater than start date.

5.2 Baseline characteristics

Baseline characteristics of participants will be presented by allocation group and follow-up 1 and 2 completion. Continuous variables will be presented including the total number, mean (SD), median (where data is not normally distributed), min, max and proportion missing. Categorical variables will be presented as total number (%) in each category. Table A1 shows the variables which will be collected at baseline for all women and their descriptive summaries.

5.3 Intention to treat

The main analyses will follow the intention to treat (ITT) principle. All participating women will be analysed by the group they were originally randomised into. Any woman randomised into the control group will be analysed as in the control group unless she requests removal from the study altogether. Furthermore, any woman randomised into the intervention group will be analysed as receiving the intervention even if she requests withdrawal from the intervention to receive standard of care. Women will be analysed as part of the pregnancy circle they were allocated to unless they join another circle prior to having attended any session in the allocated circle. This does not violate ITT since participants are still included in the analysis in the allocated treatment arm.

5.4 Withdrawals and loss to follow up

Women who choose to discontinue the group care will remain in the trial receiving standard of care and will be included in all analyses under the intention to treat principle. However, if a woman wishes to fully withdraw from the study the decision will be recorded, the withdrawal will be tabulated, and no further data will be collected. Moreover, women who withdraw before delivery will be excluded from the main analyses. However, we will still analyse their baseline data if available as stated in the protocol.

Women who have not been recorded as fully withdrawn and for whom no routine birth data is available are followed up with the Trusts by the trial team to seek their data. If unsuccessful, the participant is considered lost to follow up.

5.5 General analysis principles

The analysis of primary, secondary, and, where possible, additional outcomes will be adjusted for the stratification factors which were used during randomisation, centre, and ability to speak English, only. All main analyses will be performed on a complete case basis (ignoring missing data) for the outcomes in question. A missing category will be created for the only covariate in the models “ability to speak English”. The significance for statistical tests (alpha) will be 5%, all confidence intervals will be presented at the 95% level and all p-values will be two-sided. As the intervention arm includes group pregnancy circles, clustering will be accounted for in the intervention arm using pregnancy

circles as the units of clustering, while participants in the control arm will be modelled as a cluster size of one [1].

5.6 Risk of contamination and clustering in the control arm

Midwives leading the group antenatal sessions will potentially lead multiple different groups of women over the course of the trial. Information on midwifery group leaders is not systematically collected so that this cannot be accounted for in analysis models. Moreover, some of the midwives who have been trained to lead the group pregnancy circles (intervention) will most likely also be leading the one-on-one sessions (standard care) with women in the trial. This potentially contaminates those women in the control group receiving a one-on-one session from such a midwife, as they may unknowingly alter how they run their session. For the purposes of this study the team has not tracked which midwives are leading which sessions, so there is no way of quantifying the effect this may have. But if there is any observable impact, it will be to dilute the treatment effect, i.e. conservative. In other terms, it would be in favour of not rejecting the null hypothesis if the intervention has a positive effect on babies' health. Those risks are accepted in knowledge accordingly.

5.7 Multiple testing

No formal adjustment for multiple testing will be made. However, the number of secondary outcomes will be noted when reporting results.

5.8 Estimand Framework

Inference on the primary and secondary outcomes is complicated by the potential occurrence of intercurrent events. An intercurrent event is defined as an event that happens after randomisation which either affects the measurement, interpretation or existence of an outcome. The 2020 ICH E9 R1 addendum lays out a framework for providing a clear description of the treatment effect to be estimated from a trial (11). The table below describes the treatment effect to be estimated for REACH using the estimand framework.

Estimand framework summary for primary outcome

Objective	Primary Estimand				
	Population	Variable	Summary Measure	Treatment	Potential Intercurrent Events*
The primary objective is to demonstrate superiority of Pregnancy Circles over standard individual antenatal care on the health of babies. The primary comparison of having a healthy baby (yes/no) will be made regardless of whether women withdraw from the assigned antenatal care or receive any alternative care prior to birth in women who participate in at least one session of antenatal care.	Pregnant women registering for antenatal care with NHS maternity services excluding women defined in the exclusion list of section 6 of the trial protocol. All women who during the pandemic were not able to receive the intervention or were only able to receive a low dose of intervention are excluded.	The "healthy baby" composite consists of the following 4 components: 1. Live baby (i.e. no pregnancy loss before or after 24 completed weeks and no neonatal death within 28 days of the birth) 2. Born at term (37 weeks and above) 3. Appropriate weight for gestational age (GROW centile >9.99 & <90.01) 4. Not admitted to a neonatal care unit (which includes Intensive Care Unit, SCBU and High Dependency Unit, but NOT transitional care) A baby is considered a "healthy baby" only if the answer to all above questions is "yes", otherwise "no" .	Marginal Odds ratio and risk difference.	Pregnancy circles groups composed of 8 to 12 women during 8 sessions of 2 hours from week 16 of pregnancy (roughly 1 session every 17.5 days).	Woman stops intervention at any point for any reason Develop complications and become higher risk – woman ask to have 1-2-1 Change of service because of pandemic or relocation Rates of induction going up Woman switches circle before attending first antenatal care session Woman switches circle after attending first antenatal care session

**The treatment policy strategy (regardless of intercurrent event) will be used for all intercurrent events except for "Woman switches circle before attending first antenatal care session" where we will use the as-treated strategy (data will be analysed considering the circle women actually attended from the beginning).*

5.9 Analysis of primary, secondary and additional outcomes

The following table shows the variable type of outcomes in this study:

Outcomes:		
Binary	Continuous	Categorical
-healthy baby composite -spontaneous vaginal delivery -breast feeding initiation -breastfeeding continuation -postnatal depression -live baby -born at term -appropriate weight for gestational age -not admitted to a neonatal care unit	-attendance at antenatal care -women's empowerment -social support -self efficacy -prenatal stress -health literacy -mental wellbeing -postnatal symptoms -satisfaction with care -Infant birth weight in grams	-continuity of care -Caesarean delivery -place of birth - additional measures of satisfaction - choice and involvement in care - preparedness for labour - confidence in caring for baby - immunisation

Main analysis model (binary outcome)

The primary outcome data for the 'Healthy Baby' composite will be extracted via a postpartum maternity records audit and be analysed using a nested multilevel logistic mixed effects model with two random intercepts estimating a cluster and site-specific effects in both arms (1). Although site is a stratification factor, it has been included as a random effect due to the high number of sites included. In the intervention arm, within Pregnancy Circle cluster correlation will be accounted for and in the control arm each participant will be modelled as a cluster of size 1. The other stratification factor used in this trial is the ability to speak English and will be included as a fixed effect covariate. The resulting model will produce an odds ratio for the odds of giving birth to a 'healthy baby' in the intervention versus the control arm. The outcome will be assessed at the participant's level (mother) following this rule: if a mother has multiple births for the same delivery (e.g. twins), the 'Healthy Baby' composite will be equal to 'yes' only if all babies are healthy. Otherwise, Healthy baby will be equal to 'no'.

Specifically, let *y* be the binary outcome, *i* is the individual participant indicator (mother), *j* is the pregnancy circle indicator, *t* is the intervention indicator (0 = control, 1 = intervention), θ is the intervention effect, β_0 is an intercept term, *l* is the site indicator, and β_{ki} represents further covariates (ability to speak English in this instance). Then,

$$Logit[Pr(y_{ijl} = 1)] = \beta_0 + \theta t_i + \beta_{ki} + v_l + u_j$$

where $u_j \sim N(0, \sigma^2_u)$ is a random-effects term representing between-cluster (pregnancy circle) variation in the clustered intervention arm and $v_l \sim N(0,\sigma^2_w)$ is the random effect representing between-site variation.

Accounting for non-collapsibility

Odds ratios present the characteristic of non-collapsibility when adjusted on covariates. To deal with that phenomenon, the g-computation estimator will be used for covariate adjustment following the 6 steps described below as described in the FDA's guidance for covariate adjustment [13]:

(1) Fit a logistic model with maximum likelihood that regresses the outcome on allocation assignments and prespecified baseline covariates. The model will include an intercept term.
(2) For each subject, regardless of allocation assignment, compute the model-based prediction of the probability of response under pregnancy circle group using the subject's specific baseline covariates.
(3) Estimate the average response under pregnancy circle group by averaging (across all subjects in the trial) the probabilities estimated in Step 2.
(4) For each subject, regardless of allocation assignment, compute the model-based prediction of the probability of response under usual care using the subject's specific baseline covariates.
(5) Estimate the average response under usual care by averaging (across all subjects in the trial) the probabilities estimated in Step 4.

The estimates of average responses rates in the two allocation groups from Steps 3 and 5 will be used to estimate an unconditional intervention effect. Risk differences, and odds ratios will both be presented for primary and secondary outcomes. Confidence intervals will then be estimated using bootstrap.

Any of the secondary or additional outcomes which provide binary responses (spontaneous vaginal delivery, infant low birthweight, etc.) will be analysed using the same nested multilevel logistic mixed effects model.

Analysis of continuous outcomes

Secondary and additional outcomes providing continuous responses, such as a total score on any of the scales being used, will be analysed by assessing difference of means between intervention and control groups. This will be analysed using a partially nested mixed-effects model with heteroskedastic error terms with the Satterthwaite approximation for degrees of freedom to avoid upward bias of the type-I error rate (2). The English level will again be included as covariate in these models. The resulting model will estimate a difference of means between the intervention and control arms of the study. Specifically, let *y* be the continuous outcome, *i* is the mother participant indicator, *j* is the pregnancy indicator, *t* is the intervention indicator (0 = usual care, 1 = pregnancy circle), θ is the intervention effect, β_0 is an intercept term, and β_k represents fixed effects for English level. Then,

$$y_{ij} = \beta_0 + \theta t_{ij} + \beta_k + u_j t_{ij} + r_{ij}(1 - t_{ij}) + \epsilon_{ij} t_{ij} + v_l$$

where $u_j \sim N(0,\sigma^2_u)$ is a random-effects term representing between-cluster (pregnancy circle) variation in the clustered intervention arm, $r_{ij} \sim N(0,\sigma^2_r)$ represents individual-level variation in the

non-clustered control arm, $\epsilon_{ij} \sim N(0,\sigma^2_\epsilon)$ represents individual-level variation in the clustered intervention arm and $v_l \sim N(0,\sigma^2_w)$ is the random effect representing between-site variation.

Analysis of categorical outcomes

For categorical outcomes, proportions by intervention group and test of between group difference without covariate adjustment will be presented.

Strategy for analysis of primary outcome if model fails to converge

In case the analysis of the primary outcome described above fails to converge, the following sequential strategy will be employed (starting at 1).

	Change from previous strategy
1	Try an alternative estimation algorithm, such as the QR decomposition in meqrlogit, rather than Stata's default melogit function.
2	Change the number of integration (quadrature) points for all levels using intpoints() command
3	Try an alternative integration method other than mean–variance adaptive Gauss–Hermite quadrature like Laplace for example.
4	Remove covariate English speaking level from the model.
5	Analyse using model, not accounting for clustering within intervention groups.
6	In addition to removing random effect in (5) remove random effect for site

For any of the binary secondary outcomes that fail to converge, the process above will be repeated for the multilevel logistic mixed effects regression models.

Strategy for analysis of continuous outcomes if model fails to converge

In case the analysis of any secondary outcome fails to converge, the following strategy will be employed.

	Change from previous strategy
1	Try an alternative optimisation algorithm, such as the Newton Raphson algorithm, rather than Stata's default for xtmixed.
2	Try an alternative estimation method other than REML such as MLE or quasi likelihood based methods.
3	Fit an alternative clustering model with participants in the control arm treated as clusters of size 1.
4	Remove fixed effect covariate from the model.

5	Analyse using model, not accounting for clustering by intervention group.
6	Analyse using fixed effects model removing random effect for site in addition to random effect I (5)

Furthermore, if any issues arise for categorical additional outcomes, the outcome categories will be recoded as described in the table below.

Categorical outcomes breakdown		
Outcome	Categories	Modified categories for convergence issues
Continuity of care	1-2 midwives 3 midwives 4+ midwives	1-2 midwives 3+ midwives
Caesarean delivery	Planned Emergency None	Any Caesarean None
Place of birth	Hospital obstetric unit Hospital alongside midwifery unit Freestanding midwifery unit Home Other	Hospital (anywhere) Freestanding midwifery unit Other

If problems with other categorical outcomes names in 3.4 occur a similar collapsing strategy will be employed.

5.10 Missing Data

For the analysis of the primary and secondary outcomes we assume that the data are missing completely at random. Sensitivity analysis to this assumption will be explored in sensitivity analyses for the primary outcome using multiple imputation (6). Further details are provided in the sensitivity analysis section. Moreover, we will only consider the primary outcome for women in the situation where all of its components are non-missing (i.e. available complete-case analysis).

The primary outcome is comprised of routine birthing data. The study team's commitment to attempt to locate and follow-up women with missing data has contributed to decrease this proportion as much as possible.

5.11 Interim analyses

There are no planned interim analyses that would question the continuation of the trial. If an unplanned interim analysis should be conducted, it would be described in a separate document REACH WP3_IAP. Interim data reports have been provided to the Data Monitoring and Ethics Committee without formal stopping rules in place.

5.12 Subgroup analyses

Further analysis of the primary and secondary outcomes will be performed for the following subgroups:

a) Ethnicity.
b) Women receiving intervention prior to the Covid-19 lockdown vs women receiving intervention after lockdown (March 18th 2020).
c) Vulnerability as defined by the presence of any of the factors below (further information on the definition of vulnerability can be found in the appendices). We will also test the interaction with a vulnerability index which will be made up as the sum of the same criteria:
   - **Age** – those participants under 20 (Baseline Questionnaire – 16-19 years)
   - **Ethnicity** – any participant NOT identifying as White-British, White-Irish or White-Other (Baseline Questionnaire) (See Table 2 for full list of categories).
   - **Deprivation** – any participant living in a postcode falling in the most deprived areas in England - measured through Index of Multiple Deprivation by participant postcode (found on the Participant Information Form).
   - **Limited English Proficiency** – any participant who indicated that they do not have any English language proficiency OR do not speak English well ('not well'/'not any' categories - Baseline Questionnaire) OR if 'need an interpreter' is ticked 'yes' on the Participant Information Sheet.
   - **Social Complexity** – those participants who have been classified at booking as having 'intermediate' or 'intensive' social risk factors (Routine Data – Social Risk Profile). Social risk factors include lifestyle issues (alcohol use; substance use); recent migrant (<12 months); Refugee/asylum seeker; can't speak English; under 20 years old; domestic violence).

In order to perform the subgroup analyses, the same models for the main analyses will be used but will include an interaction term between the outcome of interest and the subgroup in question If the resulting tests for interaction are significant, then we will consider differences between individual subgroups. The purpose of the subgroup analyses will be solely for hypothesis generation due to the potentially low power of the tests for interaction. For each subgroup analysis we will report the numbers in each subgroup, summary statistics by subgroup, treatment estimates with 95% confidence intervals for each subgroup, and a p-value for the test of interaction (t-test or likelihood-ratio test).

5.13 Complier Average Causal Effect (CACE) analysis

To further assess the effect of the primary outcome healthy baby composite, the intention to treat main analysis will be compared with a Complier Average Causal Effect (CACE) analysis estimating a 'per-protocol' treatment effect. The CACE analysis will repeat the primary outcome analysis using only those participants who complied. Intervention compliance is defined as attending **three** or more antenatal circles. Further information on the definition of compliance can be found in appendix 8.2. CACE estimates are used to build upon causal modelling frameworks to yield causal estimates of the effects of intervention for individuals who comply with treatment (in our case group antenatal sessions) compared to those who would have complied in the control group. CACE effect estimates for compliance, as opposed to intention to treat, will be generated using two-stage least squares (i.e. method of instrumental variable).

5.14 Dose response analysis

We will explore the dose-response relationship between the number of sessions attended and the primary outcome in the intervention arm. The scale of this attendance variable is 0-8 where 0 = no antenatal circles attended and 8 = 8 antenatal circles attended. Dose will be used as a continuous variable assessed in a model fitted on intervention group participants only.

5.15 Protocol deviations

Major protocol deviations detailed in the appendix table such as those randomised under the incorrect stratification factor, randomised in error or received the incorrect allocation will be summarised by allocation group. For the purposes of primary analysis, participants will be analysed as they were randomised and the sensitivity of the primary outcome analysis to these assumptions will be explored through sensitivity analyses.
At the time of writing, we are aware of the following protocol deviations:

- Randomisation under incorrect stratification factor
- Incorrect allocation received (controls in intervention)
- Switched circles after first session attended

5.16 Sensitivity analyses

5.16.1 Imputation analyses assessing the uncertainty around the primary outcome analysis estimate

To assess the extent to which study results are affected by missing data, a sensitivity analysis will be performed on imputed data for primary clinical outcome and stratification variables plus baseline characteristics: age, ethnicity, main language and educational level (2). The proportion of missing values for each variable will be assessed using numerical summaries. Univariable associations between missing values of each variable and observed values of other variables will be examined to understand how reliably a missing value might be imputed (3). This will be performed by constructing separate logistic regression models after creating a binary indicator variable for each variable with missing values coded as “1” and non-missing values coded as “0”. The most applicable missing data mechanism will be informed by clinical knowledge of independent and dependent variables, reasons for missingness, and relationships between missingness and the observed values of collected variables.

Multivariate Imputation using Chained Equations (MICE) will be used to impute missing data under the expectation that both independent and dependent variables will have missing values and the data will not be monotonic missing (4). MICE replaces missing values with a random sample of plausible, imputed values drawn from their predictive distribution (5). First, an ‘imputation’ step will be performed, which involves constructing an imputation model that replaces missing data with one set of plausible values. Assuming that missing data are ‘Missing At Random’, the imputation model will specify a conditional distribution for missing values of each variable given the observed values of other variables. This imputation model will repeatedly replace missing values with a random sample of plausible values, creating a completed dataset with each imputation. The number of imputations (and thus completed datasets generated) will mirror the proportion of participants with at least one

missing value. For example, 25 complete datasets will be generated if 25% of study participants have at least one missing value (6).

A logistic regression model will be used for missing values of binary variables and a multinomial logistic regression model will be selected for missing values of categorical variables with three or more unordered categories. Missing values of categorical variables with three or more ordered categories will be modelled using ordinal logistic regression and a linear regression model will be specified for continuous variables with missing data. Auxiliary variables – that is, variables that are not included in the intended analysis of imputed variables but are the highly correlated with the imputed variables (or its missingness) – will be included in the imputation model (6).

Next, an 'estimation' step will be conducted, whereby specified analyses – as described in sections 5.7 – will be performed separately for each completed dataset that is generated during the imputation step. Finally, a 'pooling' step will be performed, whereby point estimates (e.g., sample means) and measures of precision (e.g., standard deviations) estimated in each dataset will be aggregated using Rubin's Rules to create a final estimate that accounts for between- and within-imputation uncertainty (7).

5.16.2 Primary outcome analysis following the opposite rule for multiple births

If a mother has multiple births for the same delivery (e.g. twins or triplets), the 'Healthy Baby' composite will be equal to 'yes' if at least one of the babies is healthy. Otherwise, Healthy baby will be equal to 'no'.

5.16.3 The tenability of the exclusion restriction assumption

The tenability of the exclusion restriction assumption (that the intervention effect is zero for non-compliers) in the CACE analysis will be assessed using a sensitivity analysis. Instead of restricting the intervention effect estimate amongst non-compliers to zero (as specified in the primary CACE model), we will allow the treatment effect amongst compliers AND non-compliers to be freely estimated. All other sensitivity CACE model components will be identical to the primary CACE model.

5.16.4 The effect of COVID-19 pandemic

A further analysis will be conducted to estimate the intervention effect accounting for the different phases of the COVID-19 pandemic (14). Using the primary outcome analysis described in 5.10, the effect of the trial intervention on the 'Healthy baby' criteria will be compared to the control intervention in each phase. Five phases – pre-pandemic, no dose, low/very low dose, moderate/high dose, and post-pandemic – will be used to perform a fixed-effect meta-analysis with inverse-variance weighting. The phase participants are categorise in depends on the overlap of their intervention period with pandemic periods of varying restriction and has been pre-specified (see appendix 8.3). Participants concurrently randomised into the control are categorised in the same way. Further details on the impact of covid and the definition of the dose received can be found in the appendices.

5.16.5 Sensitivity analysis of the primary outcome for protocol deviations

In this analysis, we will compare the ITT intervention effect with the ones in conditions : 1) with stratification factors correctly aligned, 2) treatment analysed as that received rather than allocated and 3) removing participants who have been randomised in error.

5.16.6 Sensitivity analysis of the dose-response relationship

In this analysis, we will change the assumptions regarding missing data in two steps: 1) we will consider that women attended all sessions where there is a missing value 2) we will impute the median and then mean value for women where the number of session(s) is missing.

6. Other analyses, data summaries, and graphs

6.1 Safety analyses

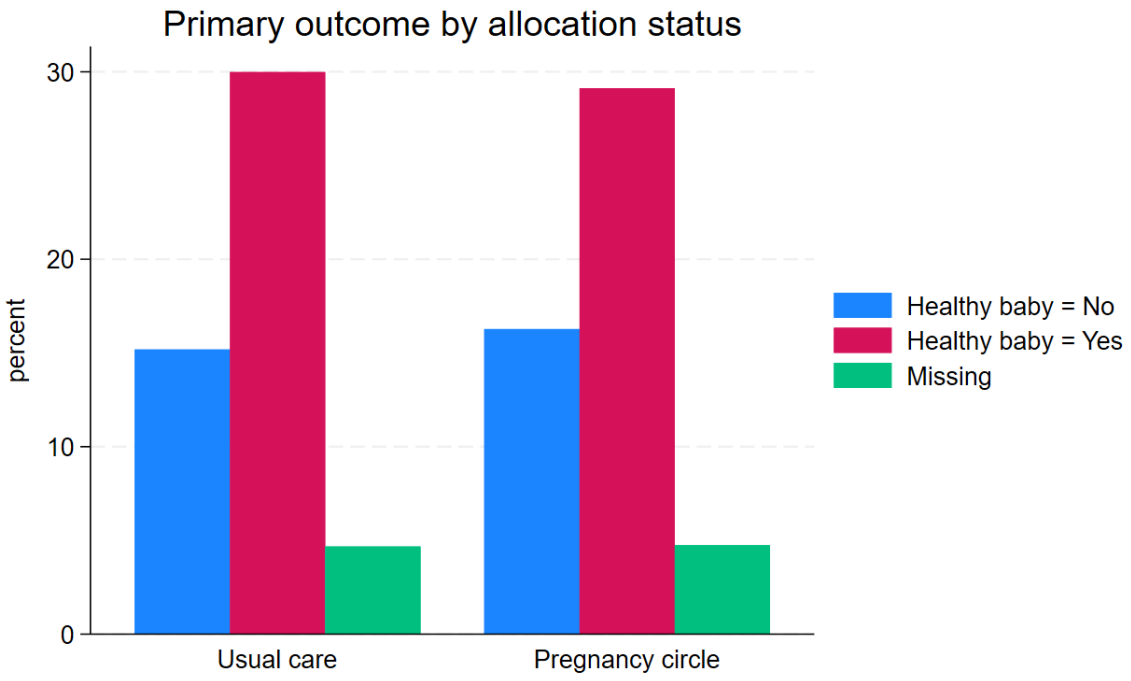
The total number (%) of serious adverse events (SAE) potentially related to the pregnancy circles intervention will be reported. Furthermore the total number (%) of SAEs, adverse events, adverse events leading to withdrawal, and the number of patients with at least one SAE will be reported by treatment group and by site.

Details on what constitutes a (serious) adverse event can be found in the study protocol v9.0.

6.2 Graphs

1) Recruitment graph over time (as example below):

2) Primary outcome graph (dummy)



3) Forest plot for sensitivity analysis 5.16.4

7. References

1. **Leahy-Warren P, Nieuwenhuijze M.** Measuring women's empowerment during the perinatal period in high income countries: A scoping review of instruments used. Heliyon. 2023 Mar 24;9(4):e14591. doi: 10.1016/j.heliyon.2023.e14591. PMID: 37064454; PMCID: PMC10102201.
2. **Parkerson GR Jr, Broadhead WE, Tse CK.** Validation of the Duke Social Support and Stress Scale. Fam Med. 1991 Jul;23(5):357-60. PMID: 1884930.
3. **Roberts, C. Batistatou, E. and Roberts, S.** Design and analysis of trials with a partially nested design and a binary outcome measure. *Statistics in Medicine.* 2016. Vol 35 pp.1616-1636
4. **Candlish, J., Teare, M.D., Dimairo, M. et al.** Appropriate statistical methods for analysing partially nested randomised controlled trials with continuous outcomes: a simulation study. BMC Med Res Methodol 18, 105 (2018).
5. **ThabaneL, Mbuagbaw L, Zhang S, et al.** A tutorial on sensitivity analyses in clinical trials: the what, why, when, and how. BMC Medical Research Methodology. 2013; 13: 92-103.
6. **Steyerberg EW.** Dealing with missing values. In: Steyerberg EW, editor. Clinical Prediction Models: A Practical Approach to Development, Validation, and Updating. New York, NY: Springer; 2009. pg 115-37 (Statistics for Biology and Health). Available from: https://doi.org/10.1007/978-0-387-77244-8_7.
7. **Jakobsen KC, Gluud C, Wetterslev J, Winkel P.** When and how should multiple imputation be used for handling missing data in randomised clinical trials – a practical guide with flowcharts. 2017; 17: 162-71. https://doi.org/10.1186/s12874-017-0442-1.
8. **Azur MJ, Stuart EA, Frangakis C, Leaf PJ.** Multiple imputation by chained equations: what is it and how does it work? Int J Methods Psychiatr Res. 201; Mar;20(1):40-9.
9. **Austin PC, White IR, Lee DS, van Buuren S.** Missing Data in Clinical Research: A Tutorial on Multiple Imputation. Can J Cardiol. 2021; 37: 1322-1331.
10. **Rubin DB.** Inference and Missing Data. Biometrika. 1976; 63(3):581-92.
11. European Medicines Agency. ICH E9 (R1) addendum on estimands and sensitivity analysis in clinical trials to the guideline on statistical principles for clinical trials. 2020 [Available from: https://www.ema.europa.eu/en/documents/scientific-guideline/ich-e9-r1-addendum-estimands-sensitivity-analysis-clinical-trials-guideline-statistical-principles_en.pdf
12. **Wiggins, M., Sawtell, M., Wiseman, O., McCourt, C., Eldridge, S., Hunter, R. Bordea, E., Mustard, C., Hanafiah, A., Hatherall, B., Holmes, V., Mehay, A., Robinson, H., Salisbury, C., Sweeney, L., Mondeh, K., Harden, A.** (2020). Group antenatal care (Pregnancy Circles) for diverse and disadvantaged women: study protocol for a randomised controlled trial with integral process and economic evaluations. BMC Health Services Research, 20(1). doi:10.1186/s12913-020-05751-z
13. FDA. Adjusting for Covariates in Randomized Clinical Trials for Drugs and Biological Products. Guidance for Industry. https://www.fda.gov/media/148910/download, 2023. Last checked: 2024- 01-05
14. **Cro, S., Morris, T.P., Kahan, B.C. et al.** A four-step strategy for handling missing outcome data in randomised trials affected by a pandemic. BMC Med Res Methodol 20, 208 (2020). https://doi.org/10.1186/s12874-020-01089-6

8. Appendices

8.1 Definition of vulnerability

Summary of evidence for vulnerability factors and adverse outcomes

Item	What we collect	Evidence of outcomes	Clinically significant difference	Hypothesis
Ethnicity	**BASELINE QU:** We used ONS categories: 1. White 2. Mixed / Multiple ethnic groups 3. Asian / Asian British 4. Black / African / Caribbean / Black British 5. Other ethnic groups NB – Can be complex: confounding of distinct concepts such as ethnicity, nationality and race	Ethnicity has been shown to be an independent risk factor for poor maternal and neonatal outcomes: **MBRRACE 2021**: Black x4, Asian x2, Mixed x2 – higher risk of dying in pregnancy **Yangmei et al (2019)** 2006-12 4.5m births – ethnicity = risk factor for preterm birth independent of country of birth (& non-white babies more likely to be more pre-term):	Background risk 8.8:100,000 Overall rate: 5.6% White – 5.5% White other – 4.6% Black Carribean 8.2% (OR 1.52) Black African – OR 1.13 South Asian – 6-6.3% (OR 1.09)	PC are a better model of AN care compared to traditional: • More time • More trust • More AN continuity • More disclosure • More education • Empowerment Evidence: Sandall/Cochrane (2016) Mw-led CoC models 24% less preterm birth 16% less likely to lose baby Rayment Jones et al (2021a; 2020) Mw-led CoC/community settings are especially effective for women with social risk factors, in particular pre-term birth & low birthweight Cohort studies suggest that PC is especially effective for 'vulnerable' groups: Byerley & Haas 2017 Carter et al. 2016
Deprivation Definition of IMD - The English Indices of Deprivation 2019 (publishing.service.gov.uk) PowerPoint Presentation (publishing.service.gov.uk)	**Participant Information Sheet** Postcode which can give us IMD **BASELINE QU:** **Tenancy** (rented council; temporary accommodation) Challenging: IMD is based on local data about employment, benefits, crime, housing health,	**MBRRACE 2021 / IMD:** Most deprived quintile x2 increased risk of dying compared to least deprived quintile **MBRRACE 2019 PERINATAL MORTALITY** IMD: Most deprived almost x2 compared to least deprived quintile; ETHNICITY: Stillbirth x1.5 for Asian & x2 for Black compared to White. **NHS Long-Term Plan**: Focus CoC on women of BAME ethnicity and from 'deprived backgrounds'	8:100,000 v 14:100,000 2.7:1000 v 1.2:1000 3.22:1000 (White) v 5..05:1000 (Asian) 7.23:1000 (Black)	

	education and accessibility of services.			
Age	**BASELINE QU:** Age (16-19, 20-25, 26-35, over 36). **NB** we did not recruit many below 20 as they were generally cared for by 'young people' teams	**MBRRACE 2019 PERINATAL MORTALITY** AGE: Mothers under 20 yrs & over 35 yrs at higher risk of perinatal mortality		
Limited English Proficiency (LEP)	**PURPLE SHEET Language -** 'do you need an interpreter' **BASELINE QU: Language** (speak Eng 'not well' or 'not any') .	**Language:** LEP affects about 9% of the population and increases the risk of perinatal mental health outcomes, low birthweight and preterm birth	(Rayment Jones et al 2021b) 2011 census suggests that circa 9% of population in London report speaking English 'not well' or 'at all' (Heslehurst et al 2018) OR 1.42 risk of low birthweight for migrant women in Europe; OR 0.24 increased risk of perterm birth; Range of OR 1.6-1.9 risk of perinatal mental health problems for migrant women. Most commonly reported risk factor for poor outcomes	

			was difficulty with language.	
Education	**BASELINE QU:** **Education** (none, GCSE, vocational, A-level, Uni, post-graduate) Measures of education complex to identify – Conelly et al 2016	**Education:** In Italy lower education was associated with worse neonatal outcomes Link with health outcomes In developing countries lower education was correlated with higher maternal morbidity.	Cantarutti et al (2017) See table below. Low – up to 8 yrs education, intermediate = years, high = 14+ years education. Raghupathi (2020) (Karlsen et al 2011) x2 the risk for women with up to 6 years (v 12 years) education.	
Social complexity / intersectionality	**Routine Data: 'Complex Social Factors' (one of:** alcohol use; substance use; recent migrant (<12 months); Refugee/asylum seeker; Can't speak English; Under 20 years old; DV) **NB:** not many women scored as 'high' in our study as these may have been given care by 'vulnerable teams' instead.	**MBRRACE 19: 'Constellation of bias'** (inc. mental health, DV, born outside UK, ethnicity, no English, living in deprived areas, unemployed, undocumented, late booker) **MBRRACE 21: 'Multiple adversity'** **MBRRACE 19 Perinatal**: Combination of age/ethnicity/deprivation much higher risk	**MBRRACE 19:** 90% of women who died had multiple problems **MBRRACE 21**: *improvements in care might have made a difference in outcome for 67% of women who died by suicide, 29% who died from substance misuse and 18% of those who died by homicide* **MBRRACE 19/perinatal** – under 25, over 35 & Black or Asian & most deprived x5 risk of neonatal mortality compared to white, 25-35 & least	

			deprived (1.21:1000 v 10.71:1000)	

8.2 Definition of compliance

There are eight pregnancy circles antenatal sessions which are at approximately the following weeks of pregnancy: 16, 25, 28, 31, 34, 36, 38, 40. Theoretically participants have the chance to attend all of these eight sessions plus one session postnatally. However, because we wanted to capture those who initiated antenatal care after the first 12 weeks, some participants may miss the first session. Some women may also miss later sessions due to delivering early (WHO define moderate to late pre-term birth as 32 to 37 weeks meaning that women could miss up to four sessions). These factors have been taken into account when defining intervention compliance as in some cases participants may only have the opportunity to take part in three antenatal pregnancy circles sessions.

Intervention compliance is therefore defined as attending three or more antenatal circles. Our rationale for choosing three or more as the minimum (rather than 1 or 2 or 4 or more) is set out below.

Number of antenatal pregnancy circles sessions	**Rationale for setting/not setting this number as cut-off point for compliance**
1-2	Attending one or two sessions only means that most of their care will have been the same as standard one to one care. One or two sessions will not have been enough to establish friendships, benefit from midwife continuity, or participate in women led discussions. They may have learned how to self-check and be part of the circle's WhatsApp group if set up.
3	***Reasonable to assume that 3 sessions is enough for women to understand (and be impacted by) the model, including self-checking; meeting other women; getting to know midwives; benefitting from woman-led discussion. Three sessions is less than half the antenatal visits for a primipara but for multipara this would represent half their antenatal visits (they have 6 antenatal appointments in standard care).***
4	Highest confidence that participants would have received benefit and true experience of intervention: this represents half or more of antenatal follow-up appointments. Sufficient time for relationship-building. BUT Will exclude those participants who booked late and delivered early as non-compliers.

	NB: In some developing countries, group care only offers 4 antenatal sessions so we can be confident that 4 sessions is enough to say they've received the intervention.

8.3 Impact of the pandemic on the Pregnancy Circles trial and dose definition

a) Overview

In line with government and NIHR guidance, we paused all recruitment for the Pregnancy Circles trial from the 18th March 2020. At this point we had recruited 1624 from our target of 1732. NHS Trusts suspended all in person group based activities and women in the Pregnancy Circles were offered one to one care. We encouraged sites to consider virtual options to continue the Circles and presented a number of other options to continue care that would be underpinned by the values of the Pregnancy Circles model (Box 1).

Options for continuing Pregnancy Circles during the pandemic

- Encourage women to continue peer support through their What's App group.
- Continuing to facilitate the groups virtually as per the schedule in the manual, using a separate clinical WhatsApp group or other technology (ask us if you would like some suggestions). Without the self-checks and palpations, session length would be reduced. The 'core values' of Pregnancy Circles (see below) can be extended to virtual meetings)
- Any virtual facilitation could be done by a single midwife, and this could be done by midwives who have to self-isolate but could still work from home.
- Pregnancy Circle women have learned how to self-monitor, and this may offer services an opportunity in providing more virtual group care and fewer 1-1:
    - Could you provide the women with a supply of urine sticks for use at home?
    - Could you support women who choose to buy home blood pressure machines to send you regular readings?
- If virtual group care is not a possibility, consider whether continuity from one of the facilitating midwives could be extended in 1-1 appointments.

The intervention continued as far as possible and we undertook several activities to support sites: held regular virtual meetings with local PIs, facilitating and research midwives; provided tailored support to challenges encountered within sites (e.g. technology, structure of virtual groups, group activities); developed and provided training in running virtual Pregnancy Circles; and gathered data on the extent to which elements of Pregnancy Circles have continued to run during the pandemic. A poster was presented at the International Normal Birth conference on how the intervention model was adapted for the pandemic consistent with its underpinning core values (see Appendix B). Alongside this, we have continued to follow-up women at 35 weeks and at three months postnatal through our follow-up questionnaires and have initiated the data extraction process to collect routinely collected data on our primary and other outcomes.

b) Pregnancy circle 'dose' offered

We have analysed information collected up until Sep 2020 on the extent to which women were offered a full 'dose' of the Pregnancy Circles intervention, in either its original or COVID-19 adapted form. We developed a scoring system which we applied to each Pregnancy Circle based on: number of circle session received before lockdown; continued continuity of carer; WhatsApp group for Circle; Self-testing; Pregnancy Circle interactive sessions implemented; non pregnancy circles interactive sessions implemented.

Scoring for intervention components during lockdown was developed to follow this schema

Post lockdown scores for intervention activity (AN only)*	
Sessions before lockdown	
7-8	6
5- 6	5
3-4	4
1-2	3
No sessions	0
Continuity of carer AN	(only counts extra if all women didn't receive this ,ie controls too)
Y	1
N (or everyone got)	0
WhatsApp group for circle	
Y	1
N	0
Self testing	
Any Y	1
N	0
PC virtual interactive AN sessions	
3+	3
1-2	2
N	0
Other non PC virtual AN session	(only counts extra if this wasn't offered to all women, ie controls too)
Y	1
N	0

Each Pregnancy Circle received a score classified as follows:

Scale of points	
6 or more points	– strong dose of intervention (or its component parts)
4-5 points	– moderate dose
2-3 points	– low dose
1 point	– very low dose
0 points	– no intervention

At the point of lockdown, 45 Pregnancy Circles had been run and completed, a further 58 circles were either already running or recruited to, and around six circles were waiting to be recruited to. Of the 58 already running or recruited to, 33 Circles have been assessed as providing a high or moderate dose of the COVID-19 adapted Pregnancy Circles during lockdown. From our target sample size of 1732 we had recruited 1624 women. We estimate that of the target sample size, 46% (n=795) received a 'full dose' of the intervention prior to the pandemic. A further 31% (n=531) had received a high or moderate 'dose' during the pandemic, 10% (n=176) had received a low or very low 'dose'; 7% (n=122) received no intervention; and 6% (n=108) were still to be recruited.

Pregnancy circle 'dose' offered

	Total groups (Pregnancy Circles)	Details of dose offered	Total participants (int + control)	Total intervention participants
Intervention as planned	45 (estimate)		795 (46% of sample size)	395
Partial intervention during lockdown	33	*High* 12 Circles (all had met face to face – 7 for 4 or more sessions; 5 who had met face to face fewer times but and then followed up with the provision of virtual groups/whatsapp/ continuity etc) *Moderate* 21 Circles (This category was either those that had met 3-4 times face to face and then had little in lockdown (8) OR Had met 1-2 times plus had other components (10). There were also 3 Circles that had not met face to face at all, but offered virtual circles)	531 (31%)	271 (117 high 154 moderate)
Low intervention during lockdown	13	*Very low* 6 Circles (Did not meet in person or virtually, but had 1 intervention component e.g. continuity of carer only) *Low* 7 Circles (2 Circles had met once, but nothing in place after; the remaining 5 did not meet face to	176 (10%)	92 (31 very low; 61 low)

		face nor offered any virtual groups, but had continuity and one other aspect eg Whatsapp group)		
Recruited but **received no intervention**	12	12 Circles; none of these circles had met face to face and no additional intervention components offered during lockdown. (most mid recruitment)	122 (7%)	63
Not yet recruited	-	-	108 (6%)	-

8.4 Dummy result tables

Table 1: Baseline Characteristics of recruited women in REACH*

	Group		
Baseline characteristics	**Usual care (N=)**	**Pregnancy circle (N=)**	**Total (N=)**
Age (years)			
N of non-missing values			
Mean (sd) [IQR]			
Ethnicity - N (%)			
White-British			
White-Irish			
Other-White			
Black or Black British-Caribbean			
Black or Black British-African			
Black or Black British-other Black			
Asian or Asian British-Indian			
Asian or Asian British-Pakistani			
Asian or Asian British-Bangladeshi			
Asian or Asian British-Chinese			
Asian or Asian British-Other Asian			
Mixed-White & Black Caribbean			
Mixed-White & Black African			
Mixed-White & Asian, Mixed-Other			
Other - Arab, Any other ethnic group			
Missing			
What is your main language - N (%)			
English			
Other			
Missing			
How well can you speak English - N (%)			
Very well or well			

Not well or I do not speak any English Missing		
What is your highest educational qualification? - N (%) Don't have any GCSE or similar (exams after 5 years of high school) Vocational qualifications (e.g. NVQ, BTEC) A level or similar (exams after 7 years of high school) University undergraduate degree Postgraduate degree Missing		
Attendance to antenatal care - N (%) Woman didn't miss any session Missed one or two Missed 3 or more Missed all sessions Can't remember Missing		
Revised prenatal distress scale N of non-missing values Mean (sd) [Min, max] IQR		
Emotional Wellbeing (SWEMWBS) N of non-missing values Mean (sd) [Min, max] IQR		

**This table will be produced by follow up completion (35 weeks pregnant and 3 months post-partum)*

Table 2: Results for analysis of primary, secondary and additional binary outcomes

	Number included in analysis		Summary measure		Treatment effect			
Outcome	Usual care n (%)	PC* n (%)	Usual care % of yes	PC* % of yes	Odds ratio [95% CI]	Risk diff. [95% CI]	p-value	ICC** [95 % CI]***
Healthy baby								
Baby alive								
Gestation at birth >36wk								
Appropriate weight for gestational age								
Admitted to a neonatal care unit								
Spontaneous vaginal delivery without instruments								
Breast feeding initiation								
Breast feeding continuation at Month 3 pp								
Postnatal depression at Month 3 pp								

Primary outcome is in bold and underlined and secondary outcomes are in bold and italic
** PC for Pregnancy Circle group*
*** ICC for clustering in intervention group (cluster is PC group)*
**** Confidence interval is established using Swiger's method*

Table 3: Results of secondary and additional continuous outcomes at 35 weeks and 3 months post-partum

	Number included in analysis		Summary measure		Treatment effect			
Outcome	Usual care n (%)	PC n (%)	Usual care Mean (SD)	PC Mean (SD)	Mean difference	(95% CI)	p-value	ICC** [95 % CI]***
Pregnancy-related Empowerment Scale (PRES) Range 16-64 Week 35								
Friends and family test Range 1-5 Week 35 Month 3 pp								
Emotional Wellbeing (SWEMWBS)Range 7-35 Week 35								
The Duke-UNC Functional Social Support Questionnaire Range 8-40 Month 3 pp								
Pearlin Mastery Scale Range 7-28 Month 3 pp								
Revised prenatal distress scale Range 0-16 Week 35								
Infant birth weight (g) Month 3 pp								
Health literacy Range 1-20 Week 35								

Post natal symptoms (NPEU) Range 0-5 Month 3 pp								
Number of antenatal care sessions attended								

Secondary outcomes are in bold and italic
**PC is for Pregnancy Circle group*
*** ICC for clustering in intervention group (cluster is PC group)*
**** Confidence interval is established using Swiger's method*

Table 4: Analysis of categorical Additional Outcomes

	Number included in analysis		**Summary measure**		
Outcome	Usual care n (%)	PC n (%)	Usual care %	PC %	P value for between group difference
How do you feel about the care you received from midwives? (week 35)					
very happy					
fairly happy					
not very happy					
very unhappy					
Missing					
How do you feel about the care you received from midwives (before the birth of your baby)? (3 months pp)					
very happy					
fairly happy					
not very happy					
very unhappy					
Missing					
Caesarean delivery					
Planned					
Emergency					
None					
Missing					
Place of birth					
Hospital obstetric unit					
Hospital alongside midwifery unit					
Freestanding midwifery unit					
Home					
Other					
Missing					

Has your new baby had their routine immunisations? Yes 2 months No 2 months Yes 3 months No 3 months Missing			
How many midwives did you have during care? (week 35) 1-2 3 4+ Don't know Missing			
Do the midwives you saw got to know you and remembered you and your progress? (week 35) Very satisfied Quite satisfied Not at all satisfied Don't know/can't remember Missing			
Do you feel that midwives have been sensitive to your cultural and/or language needs? (week 35) Yes Definitely Yes, a little No, not at all Don't know/can't remember Missing			
Were you offered any of the following choices about where to have your baby? (week 35 & 3 months pp) A choice of different hospitals In a midwife-led unit or a birth centre In a consultant-led unit At home I was not offered any choices I was not offered any choices due to medical reasons Don't know Missing			
Were you involved enough in decisions about your care? (3 months pp) Yes Always Yes – sometimes No and I wanted to be			

No and I did not want to be Don't know Missing			
How prepared did you feel for labour and birth? (Week 35 & 3 months pp) Very well Quite well Not very well Not at all well Missing			
How well did you manage during labour? (Week 35 & 3 months pp) Very well Quite well Not very well Not at all well Missing			
How confident did you feel about caring for your baby in the first week after the birth? (3 months pp) Very confident Fairly confident Not very confident Not at all confident Don't know/can't remember Missing			
Have you received enough help and advice from a midwife and/or health visitor about your baby's health, care and progress? (3 months pp) Yes, definitely Yes to some extent No, and I wanted help/advice No, but I did not need any Don't know Missing			

Table 5. Descriptive table of Breast feeding initiation and continuation

	Usual care (N=)	Pregnancy circle (N=)	Total (N=)
Outcome	**N participants (%)**	**N participants (%)**	
Birth:			
Breast feeding method			
Breastfeeding exclusive			
Artificial exclusive			
Mixed breast and artificial			
Other			
Not applicable			
Not collected			
Missing			
First few days after the birth:			
What type of milk			
Only breastmilk			
Only formula milk			
Breast AND Formula Milk			
Missing			
Month 3 post-partum:			
What type of milk			
Only breastmilk			
Only formula milk			
Breast AND Formula Milk			
Missing			

Table 6. Protocol deviations summary

	Usual care (N=)	Pregnancy circle (N=)	Total (N=)
Randomised under incorrect stratification factor			
Switched circle after 1st session attended			
Incorrect allocation received			
Other			

Table 7: Adverse and serious adverse events

	N SAEs		N participants experiencing SAEs		N unexpected SAEs related to the intervention		N AEs		N participants experiencing AEs	
	PC	UC	PC	UC	PC	UC	PC	UC	PC	UC
Site										
ASP										
BDH										
CLT										
EST										
EPS										
HAR										
HHT										
IPS										
LGT										
QEH										
NWH										
RFH										
RLH										
SAS										
STH										
WHH										
WHX										
WOR										

**PC – Pregnancy circle, UC – Usual care*

Table 8 – Results for subgroup analysis of primary outcome

	Number included in analysis		Healthy baby - yes			
	PC* N	UC* N	PC* N (%)	UC* N (%)	OR 95%CI	p-value for interaction **
Ethnicity						
White						
British						
White-Irish						
Other-White						
Black or						
Black British-Caribbean						
Etc…						
Vulnerability						
Yes						
No						
Vulnerability index						
0						
1						
2						
3						
4						
5						

*PC – Pregnancy circle, UC – Usual care
** Likelihood ratio test when multiple categories

Table 9 – Results of Complier-Average Causal Effect analysis investigating the effect of the intervention on the Healthy baby outcome amongst compliers*

Estimator	N	OR 95%CI	RD 95%CI	p-value
ITT				
CACE				

**95%CI, 95% confidence interval; ITT, intention-to-treat; CACE, complier-average causal effect*

Table 10 – Results of dose-response relationship between primary outcome and number of antenatal care sessions attended

	Number included in analysis (N)		Regression coefficient (β)	95%CI	P-value
N of antenatal care session(s)					

Table 11. Sensitivity analysis: multiple imputation of primary outcome and spontaneous vaginal delivery

Outcome	n (%) (complete case data)	n (%) (multiply imputed data)	OR 95%CI (complete case data)	OR 95%CI (multiply imputed data)	RD 95%CI (complete case data)	RD 95%CI (multiply imputed data)	P-value (complete case data)	P-value (multiply imputed data)
Health baby								
Spontaneous vaginal delivery without instruments								

**Primary outcome is in bold and underlined*

Table 12. Sensitivity analysis: Results for primary analysis for multiple births considering 'Healthy Baby' composite equals 'yes' if at least one of the babies is healthy. Otherwise, Healthy baby equals 'no'.

	Number included in analysis		Summary measure		Treatment effect			
Outcome	Usual care n (%)	PC* n (%)	Usual care % of yes	PC* % of yes	Odds ratio [95% CI]	Risk diff. [95% CI]	p-value	ICC** [95 % CI]***
Healthy baby								

** PC for Pregnancy Circle group*

*** ICC for clustering in intervention group (cluster is PC group)*

**** Confidence interval is established using Swiger's method*

Table 13. Sensitivity analysis investigating the robustness of CACE analysis* results to the exclusion restriction assumption (that the treatment effect is zero for non-compliers)

	OR (95%CI)	RD (95%CI)	P-value
Main analysis			
Sensitivity analysis			

**effect of intervention amongst 'compliers' when exclusion criterion does/does not apply*

Table 14. Sensitivity analysis: Results of analysis accounting for the COVID-19 pandemic

Pandemic phase	Number analysed N*	OR (95%CI)	RD (95%CI)	P-value
Pre-pandemic				
No dose				
Low/very low dose				
Medium/high dose				
Post-pandemic				
Pooled, aggregate effect				

Table 15. Sensitivity analysis: Results of dose-response relationship between primary outcome and number of antenatal care sessions attended with different missing data assumptions

Missing data assumption	Number included in analysis (N)	Regression coefficient (β)	95%CI	P-value	
N of antenatal care session(s) (original assumption)					
N of antenatal care session(s) (mean imputation)					
N of antenatal care session(s) (median imputation)					
N of antenatal care session(s) (attendance imputation)					

Table 16 Sensitivity analysis for protocol deviations

Protocol deviation	Number analysed N(%)		OR (95%CI)	RD (95%CI)	P-value
	PC*	UC*			
Randomised under incorrect stratification factor					
Switched circle after 1st session attended					
Incorrect allocation received					
Other					

**PC: pregnancy circle, UC: usual care*