Statistical Analysis Plan

Final Version 2.1

Trial Title	**A Multicentre Randomised Controlled Trial of Induced Endometrial Scratch in Women Undergoing First Time In Vitro Fertilisation**
Short Title	**The Endometrial Scratch Trial**
Version Number	**2.1**
Funder and Reference	**NIHR HTA 14/08/45**
Trial registry / registration number	**ISRCTN registry / ISRCTN23800982**
REC	**16/SC/0151**
Sponsor	**Sheffield Teaching Hospitals NHS Foundation Trust**

Authored by

Dr Munya Dimairo
Trial Statistician
Sheffield CTRU, University of Sheffield

22/07/2020
Date

Approvers

Prof Stephen Walters
Senior Trial Statistician
ScHARR, University of Sheffield

22/07/2020
Date

Mr Mostafa Metwally
Chief Investigator
Sheffield Teaching Hospital

28/07/2020
Date

Dr Alan Watkins
TSC Senior Statistician
Swansea University

22/07/2020
Date

List of Abbreviations

AE	Adverse Event
AIR	Adjusted Incidence Rate
AIRR	Adjusted Incidence Rate Ratio
AMH	Anti-mullerian hormone
BMI	body mass index
CC	Complete Case
CI	Confidence Interval
CONSORT	Consolidated Standards Of Reporting Trials
CRF	Case Report Form
CTRU	Clinical Trials Research Unit
eCRF	electronic Case Report Form
ES	Endometrial Scratch
DMEC	Data Monitoring and Ethics Committee
DMG	Data Management Group
FSH	follicle stimulating hormone
GLM	generalised linear model
hCG	human chorionic gonadotropin
HFEA	Human Fertilisation and Embryology Authority
HTA	Health Technology Assessment
ICH	International Conference of Harmonisation
ICM	Inner cell mass
ICSI	Intracytoplasmic Sperm Injection
IQR	interquartile range
IR	Incidence Rate
IRR	Incidence Rate Ratio
ITT	Intention-to-treat
IUI	Intrauterine insemination
IVF	In Vitro Fertilisation
LBR	Live birth rate
LTFU	lost to follow-up
Max	Maximum
MLE	maximum likelihood estimate
Min	Minimum

NEQAS	National External Quality Assessment Service
NHS	National Health Service
NIHR	National Institute for Health Research
OR	Odds Ratio
PP	Per-Protocol
PUL	Pregnancy of Unknown Location
PV	Per Vaginal
QoL	Quality of Life
RCT	Randomised Controlled Trial
RD	Risk Difference
RR	Relative Risk / Risk Ratio
SAE	Serious Adverse Event
SAP	Statistical Analysis Plan
SCRAM	Sheffield CTRU Randomisation
SET	single embryo transfer
SD	standard deviation
SOP	Standard Operating Procedure
TAU	Treatment As Usual
TE	Trophectoderm
TMG	Trial Management Group
TSC	Trial Steering Committee
WHO	World Health Organization

Table of Contents

List of Tables

List of Figures

1. Introduction

This Statistical Analysis Plan (SAP) provides detailed guidance for the Trial Statistician undertaking the statistical analysis and reporting for the Endometrial Scratch (ES) trial. This section gives a brief background of the trial, the primary research question under investigation, the study design used to address the research questions, and key documents guiding the development of this SAP.

1.1 Brief background and primary research question

The ES procedure is known to improve the pregnancy rates in women undergoing assisted conception – In Vitro Fertilisation (IVF), with or without Intracytoplasmic Sperm Injection (ICSI), with a history of implantation failure. However, the effect of ES procedure in women having IVF or ICSI treatment for the first time has not been adequately investigated. This trial, therefore, aims to investigate the clinical and cost-effectiveness of the ES procedure performed in the midluteal phase prior to a first time IVF/ICSI cycle using either antagonist or long protocols on the chances of achieving clinical pregnancy and live birth. Full details of the trial background are provided in a published protocol [1].

The trial is funded by the National Institute for Health Research (NIHR) Health Technology Assessment (HTA) Programme (ref. HTA 14/08/45). The trial sponsor is the Sheffield Teaching Hospitals NHS (National Health Service) Foundation Trust. The trial is registered with the ISRCTN registry (ISRCTN23800982).

1.2 Trial design

This is a two-arm, multicentre study involving 16 Fertility Units across the UK. The study is a parallel-group, superiority, pragmatic, confirmatory, open-label, and individually randomised controlled trial (RCT). The trial compares IVF without ES versus IVF plus ES intervention arms, in women undergoing first IVF treatment. The trial was designed with an internal pilot phase with pre-planned STOP/GO criteria focusing to assess the feasibility aspects of conducting the trial. For consistency throughout this SAP, the control arm, which is the IVF without ES, shall be referred to as 'treatment as usual' (TAU) and the IVF plus ES intervention arm as ES.

1.3 Aims and objectives

As highlighted in Section 1.1, the main aim of the trial is to investigate the effect of ES procedure performed in the midluteal phase prior to a first time IVF cycle (with or without ICSI) on the chances of achieving a clinical pregnancy and live birth. The specific objectives are:

a) To conduct a trial to examine the clinical effectiveness, cost-effectiveness, and safety of the ES procedure,

b) To assess as part of an internal pilot phase:
   i. the availability of eligible participants and the feasibility of recruitment of women into the main trial including the need to translate study material into other languages,
   ii. the feasibility of scheduling the ES procedure at the correct time in the treatment pathway.

2. Documents guiding the SAP

This SAP is written in conjunction with the International Conference of Harmonisation (ICH) topic E9 on statistical principles for clinical trials [2], guideline on clinical trials SAPs [3] applicable Standard Operating Procedures (SOPs) from the Sheffield Clinical Trials Research Unit (CTRU), particularly ST001 [4], and the trial protocol [1] and related amendments.

3. The scope of the SAP

As highlighted in Section 1.3, this trial was designed with an internal pilot phase only to assess the feasibility of patient recruitment and scheduling of the ES procedure. Furthermore, there is a nested biomedical tissue sub-study investigating endometrial factors that play a role in embryo implantation. The trial was also designed with health economics evaluation to address the cost-effectiveness of the ES intervention. However, this SAP focuses on addressing the clinical effectiveness related research questions of the trial and internal pilot feasibility objectives highlighted in Section 1.3. The biomedical tissue sub-study and health economic evaluation aspects, which are out of the scope of this SAP, will be addressed elsewhere.

4. Outcomes measures and timing of assessments

This section describes the outcome measures, which are used to evaluate trial objectives relating to the internal pilot phase, clinical effectiveness, and safety of the ES intervention. The timing of the outcome measures is stated, starting with the primary outcome followed by secondary and safety outcomes. The outcomes relating to biomedical tissue sub-study and health economic evaluation are excluded, as they are out of the scope of this SAP.

4.1 Internal pilot primary outcomes

The following feasibility outcomes were assessed at the end of March 2017 to evaluate the STOP/GO criteria as guided by:

a. The average number of women recruited per site per month,
b. The percentage of women scheduled to receive their ES procedure who received it at the correct time, assessed by comparing the date the ES procedure was scheduled and the actual date the procedure took place.

The conduct of the main trial will be deemed feasible if the corresponding criteria are met:

a. At least 108 participants have been recruited before the end of December 2016, which equates to 3 participants being recruited to the study on average per site per month,
b. At least 75% of women scheduled to receive their ES procedure have received it at the correct time point.

The Trial Steering Committee (TSC) tasked to provide 'independent' oversight of the trial on behalf of the sponsor and the funder reviewed the feasibility progression criteria and provided feedback to the funder.

4.2 Primary outcome

To address the primary research question, the primary endpoint is the live birth rate (LBR) measured by the number of live births after 24 weeks gestation within the 10.5-month post egg collection follow-up period relative to the number of women randomised. Multiple live births per mother (such as twins or triplets), misclassification, and missing data will be dealt with as described in Sections 11.9.1.2 to 11.9.1.4.

4.3 Secondary outcomes

To address other secondary trial objectives, the following secondary outcome measures are recorded :

1. Acceptability of the ES procedure, as measured by;
    a. Pain rating (on a score of 0 to 10) and tolerability (yes or no) within 30 minutes of the procedure,
    b. Pain rating (on a score of 0 to 10) directly after 24 hours and 7 days post-procedure.
2. Implantation rate as measured based on a positive serum Beta-human chorionic gonadotropin (hCG) or by a positive urine pregnancy test on approximately day 14 following the egg collection;
3. Ectopic pregnancy as measured by the rate of pregnancy outside the normal uterine cavity;
4. Clinical pregnancy rate measured based on an observation of viable intrauterine pregnancy with a positive heart pulsation seen on ultrasound at/after 8 weeks gestation;
5. Miscarriage rate measured based on a spontaneous pregnancy loss, including pregnancy of unknown location (PUL) prior to 24 weeks gestation within the 10.5-month post egg collection follow-up period;
6. Multiple birth rate defined based on the birth of more than one living foetus after completed 24 weeks gestation;
7. Preterm delivery rate as measured by live birth after 24 weeks before 37 weeks gestation within the 10.5-month post egg collection follow-up period;
8. Stillbirth rate based on the delivery of a stillborn foetus showing no signs of life after 24 weeks gestation within the 10.5-month post egg collection follow-up period;
9. Details of participant's IVF cycles including the;
    a. Number of eggs retrieved,

b. Number of embryos generated 1 day after egg collection,
c. Quality of the embryos transferred measured using NEQAS and Gardners grading system,
d. Number of embryos replaced,
e. Day of embryo replacement.

4.4 Safety outcomes

Adverse events (AEs) and Serious Adverse Events (SAEs) experienced by participating women during the course of the trial are recorded. These outcomes are detailed in Section 11.14. It should be noted that the follow-up period (time to study end) for participating women is variable depending on the pregnancy outcome and other aspects. For example, the safety outcomes of women who experience a miscarriage or stillbirth are only recorded to this point due to ethical considerations as detailed in the trial protocol.

Recorded safety outcomes (AEs and SAEs) relating to born babies 6 weeks post-partum will be reported as described in Section 11.14.2.

5. Sample size estimation

The primary outcome is the LBR defined as a live birth after completed 24 weeks gestation within the 10.5-month post egg collection follow-up period. The number of women randomised to each treatment arm will be the denominator used to calculate the LBR. Available data from the Human Fertilisation and Embryology Authority (HFEA) suggests a LBR of 32.8% and 27.3% in women under 35 and aged 35 to 37, respectively. For the sample size calculation, we therefore, assume a 30% LBR in the TAU arm (control) and that a 10% absolute increase to a 40% LBR, a relative risk (RR) of 1.33, in the intervention arms is of clinical and practical importance. The proposed effect size of 10% absolute difference in LBR is large, but it is believed that an effect of such magnitude is needed to change clinical practice (there is a 5% absolute difference in LBR between women aged under 35 and 35-37) and is less than that observed in the systematic reviews [5,6], where the RR estimates ranged from 1.83 to 2.29.

To preserve at least 90% power of detecting a 10% absolute difference in LBR rates between intervention arms, as statistically significant at the 5% two-sided level, the trial would require a total of 992 women (496 per arm). In addition, we anticipate difficulties of follow-up for patients who have been referred from NHS Trusts other than the participating Fertility Unit. As a result, the trial would require a total of 1044 women (522 per arm) after adjusting for an expected follow-up dropout rate of 5%.

6. Eligibility screening

This section detail inclusion and exclusion criteria used for screening women for eligibility into the trial as described in version 6 of the protocol.

6.1 Inclusion criteria

1. Women expected to be aged between 18 and 37 years (inclusive) at time of egg collection;
2. First time IVF with or without ICSI treatment using the antagonist or long protocol only;
3. Expected to receive treatment using fresh embryos;
4. Expected good responders to treatment, with:
    a. Ovulatory menstrual cycle (Regular menstrual cycles defined by clinical judgement or with ovulatory levels of midluteal serum progesterone as defined by local laboratory protocols);
    b. Normal uterine cavity (assessed by transvaginal sonography at screening and no endometrial abnormalities such as, suspected intrauterine adhesions, uterine septa, submucosal fibroids or intramural fibroids exceeding 4 cm in diameter as assessed by the investigator that would require treatment to facilitate pregnancy);
    c. Expected good ovarian reserve [assessed clinically, biochemically (FSH, follicle stimulating hormone < 10 & normal follicular phase oestradiol levels and or normal AMH, anti-mullerian hormone), and or sonographically (antral follicle counts) and no history of previous radiotherapy or chemotherapy]. All laboratory/ultrasound standards are based on local normal reference ranges;
    d. Single embryo transfer (SET) expected.
5. Local procedures have been/will be followed to exclude relevant vaginal/uterine infections prior to starting treatment;
6. Willing to use an appropriate method of barrier contraception if randomised to ES in the cycle where the ES procedure is performed;
7. Understands/willing to comply with the protocol.

6.2 Exclusion

1. Previous trauma/surgery to the endometrium (e.g. resection of submucous fibroid, intrauterine adhesions.);
2. BMI of 35 kg/m^2 or greater;
3. Known grade 4 (severe) endometriosis;
4. Currently participating in any other fertility study involving medical/surgical intervention;
5. Expected to receive protocols other than antagonist or long (e.g. ultra-long protocol);

6. An endometrial scratch (or similar procedure, e.g. endometrial biopsy for the collection of Natural Killer Cells) is planned;
7. Previously randomised into this trial.

7. Trial features to minimise bias

This section describes design measures put in place to avoid the potential of bias in evaluating the effectiveness of the ES intervention focusing on randomisation, its concealment, blinding and masking, and the primary outcome measure. Additional measures to minimise bias during the statistical analysis such as dealing with missing data and potential misclassification issues are addressed in Section 11.9.1.2 to 11.9.1.4.

7.1 Design, randomisation, and concealment

The trial utilises objective outcome measures to evaluate research questions relating to pregnancy. These are unlikely to be affected by the placebo effect in the control arm. As a result, administering a sham ES procedure in the control arm was viewed as unnecessary.

Eligible women were randomised to either ES or TAU interventions with an equal chance of receiving the two interventions with informed consent using the web-based Sheffield CTRU Randomisation (SCRAM) system. Permuted block randomisation algorithm stratified by recruiting site (Fertility Unit) and treatment protocol (antagonist or long) was used. Random permuted blocks of variable size were used to ensure participants are allocated evenly to each arm of the trial at each site and within treatment protocol, and not to balance the number of women assigned between protocols per site. We used blinded variable block sizes documented in a restricted access folder to minimise the chances of predicting future allocation sequence by those involved in the randomisation process. The block sizes will be disclosed in the trial report after trial completion during reporting. The randomisation process and procedures were guided by the ST007 Sheffield CTRU SOP [7].

A member of the local research team logged on to the SCRAM web-based system and entered the participant's details. The participant was then allocated a participant identification number. Details entered into the system as specified in the ST007 Sheffield CTRU SOP [7] Randomisation Request Form included confirmation of signed informed consent and eligibility, recruiting site, and planned IVF protocol. Randomisation was only undertaken once the patient's IVF treatment protocol has been decided. Participants were then randomly allocated to either the ES or TAU arm of the trial using the SCRAM web-based system. A research team member who undertook randomisation documented the treatment allocation even though this information is automatically generated and retained by the SCRAM web-based system.

7.2 Blinding and trial integrity issues

This is an open-label trial. The nature of the ES intervention makes it impossible to blind the participants, clinical investigators, trial staff, and outcome assessors. However, the trial employs objective hard-endpoints relating to pregnancy outcomes, such as LBR, to address the primary research question. Thus obviating the potential of assessment bias of the primary and important secondary outcome measures.

Trial Statisticians and a Health Economist were blinded to treatment allocation during the course of the trial until the point of data freeze before any analysis. A Trial Statistician or delegated team member provided the Data Monitoring and Ethics Committee (DMEC) related summaries in a blinded manner as a default approach. However, in the case of the need to unblind statistical summaries/reports to the DMEC on their request, as guided by the DMEC Charter, a Statistician within the Sheffield CTRU but external to direct conduct of the trial or data management team was tasked to produce the relevant reports where appropriate. Furthermore, the trial SAP and related amendments were written and signed off prior to data freeze before any form of statistical analysis. The nature of information known to the Trial Statistician prior to amending the SAP (e.g. with or without knowledge of treatment allocation) is disclosed in Section 15.

8. Trial monitoring and interim analyses

The conduct of this trial was guided and monitored by three oversight committees as governed by internal Sheffield CTRU SOPs, GV001 [8], GV002 [9], and GV003 [10], trial protocol, and the DMEC Charter. The committees are the Trial Management Group (TMG), the TSC, and the DMEC.

The trial is a fixed sample size design with only one formal statistical analysis at the planned scheduled end when all participants are recruited and completed outcome assessments. Thus, there are no planned interim analyses to allow early stopping using formal statistical rules. However, the trial was independently monitored by the DMEC within the premise of the DMEC Charter, which was agreed and signed by all the members. A recommendation to stop the trial could be made by the DMEC based on safety reasons as stipulated in the DMEC Charter. In addition, there was an option for the DMEC to perform an *ad hoc* one-off futility analysis based on stochastic curtailment at their discretion when the need arises for other reasons (although unlikely to take place) . Interim DMEC reports were produced with relevant summary statistics as requested by the DMEC. Periodic interim reports were provided to the DMEC in a blinded and unblinded manner at the request of the DMEC, as highlighted in Section 7.2.

9. Data sources and data management

All the data to address the research questions are recorded on electronic Case Report Forms (eCRFs) hosted by the Sheffield CTRU. The Data Management Group (DMG) developed the CRFs in close consultation with the Trial Statisticians, Health Economists, and TMG members, as guided by a relevant DM003 internal SOP [11] to ensure that all relevant data are appropriately collected to address trial objectives. Data will be stored on the Sheffield CTRU database system, which offers restricted access to certain trial staff depending on their duties and responsibilities. The Sheffield CTRU data management unit validated and queried electronic data for inconsistencies during the course of the trial as governed by the processes and procedures stipulated in the Data Management Plan. The Trial Statistician will conduct any additional validation checks when appropriate before the data lock and sign off guided by the relevant SOPs such as DM005 [12].

10. Definition of analysis populations and subgroups

This section defines the primary analysis populations, safety population, and other secondary analysis populations, which will be used for sensitivity analyses. Protocol violations judged to be important in defining the Per-Protocol (PP) analysis population for sensitivity analysis are stated as guided by clinical advice from the TMG. Pre-specified subgroups for further exploratory effectiveness analysis of the ES intervention as stated in the protocol are outlined.

10.1 Analysis populations

The primary analysis is based on an Intention-to-treat (ITT) population as defined in Table 1. Additional sensitivity analyses as described in Section 11.9.2 will be undertaken based on PP and Complete Case (CC) populations defined in Table 1 where appropriate. Clinical input has been sought through the TMG and using relevant statistical literature [13] to help define the PP population.

Table 1: Definitions of analysis populations

Analysis population	**Patient inclusion criteria**
ITT	1. All participants allocated to either ES or TAU interventions sequence **and**, 2. Consented to take part in the study (excludes women who withdrew consent and explicitly stated that their data should not be used) **and**, 3. Treatment assignment during analyses is as allocated at randomisation regardless of what happens after randomisation.
PP [a]	A subset of the ITT who complied with the 'protocol' requirements. This excludes participants who:

	1. Failed to meet any inclusion criterion (during or after screening) as stipulated in the protocol but were included (consented and randomised) for some reason(s) or; 2. Switched randomised treatment in either direction for some reason(s) (e.g. allocated to TAU but received ES from within or outside the trial or allocated to ES but failed to receive it before IVF) or; 3. Had spontaneous pregnancy or; 4. Embryo not generated for any other reason(s), such as because of failed fertilisation or; 5. Cycle cancelled due for any other reason(s), such as because of insufficient follicle development or; 6. Failed to use contraception prior to ES and their procedure could not be rescheduled or; 7. Failed to receive treatment using fresh embryos (i.e., frozen embryo transfers are excluded) or; 8. Were randomised but failed to receive IVF for some reason(s) or; 9. Were known to have received any protocols other than the antagonist or long (e.g. ultra-long protocol). Treatment assignment during the analysis will be done as per the randomisation sequence. Of note, based on TMG advice, women who were randomised to ES but received ES procedure outside the trial (if any) will be included in this analysis.
CC	A subset of the ITT population but includes only women with outcome measurements at a specific follow-up time **and** treatment assignment during analysis as per the randomisation sequence.
Safety	Consented to take part in the study **and** treatment assignment as per intervention received and not randomised: 1. Women randomised to the ES arm but did not receive the ES procedure for some reason(s) will be assigned to the TAU arm, 2. Women randomised to the TAU arm but received ES prior to IVF for some reason(s) will be assigned to ES arm, 3. Women who fail to receive any of the interventions will be excluded.

[a] stated exclusions, which are part of 'protocol violations' are captured on eCRFs. Thus, no adjudication by the members of the research team is required to define these exclusions; ITT: Intention-to-treat; CC: Complete case. For safety, sensitivity analysis will be based on treatment assignment as randomised.

10.2 Prespecified subgroups

As stipulated in the protocol, subgroup analyses will be performed for the primary outcome (Section 4.2) and secondary outcomes (items 2 to 8 in Section 4.3). The objective is to explore subgroups of women who are more likely to benefit from the ES intervention. The following six subgroups of interests for exploratory analyses have been pre-specified based on clinical input:

1) Day of embryo transfer (2, 3, 4, 5, or 6),
2) Fertilisation method (IVF, ICSI or split ICSI). There is a possibility for some women to receive a split ICSI,
3) Type of protocol (long treatment or antagonist). The effect of down-regulation mechanisms (day 2 or 21) under the long treatment protocol is deemed irrelevant based on clinical advice so these will be combined,
4) Embryo transfer (single or double),
5) Nature of embryo used (frozen or fresh),
6) History of miscarriages (0-2 or $\geq$3),
7) Cycle programming (yes/no),
8) Delay of IVF after ES procedure (ES arm only).

Details of statistical methods to undertake subgroup analyses and reporting are described in Section 11.11. It should be noted that the delay of IVF treatment after ES procedure, which only occurs in the ES arm, will be described in Section 11.7 and further explored in Section 11.12.

11. Outline of statistical analyses

This section outlines the statistical analyses framework to be adopted, beginning with how trial data and results will be reported. The description of the statistical methods used to analyse outcomes to address trial research questions is provided in order of importance, starting with the primary outcome then followed by secondary and safety outcomes. Dummy tables and figures of results are provided only to guide the Trial Statistician(s) during analysis and reporting.

11.1 Reporting framework of trial data

Since this study is a two-arm, parallel-group, multicentre RCT, the analysis of trial data and reporting will be guided by the revised CONSORT statement for parallel-group individually randomised trials [14,15]. A detailed CONSORT flow diagram from screening to the end of the trial will be constructed using the information summarised in Section 11.2 at the discretion of the Trial Statistician (e.g. in line with the preference of the target journal).

Baseline summary statistics will be reported by treatment arm (ES or TAU) and overall. Comparability between treatment arms (at randomisation or before interventions) will be descriptively reported without any statistical significance testing [16–18]. Any observed differences in baseline characteristics and demographics believed to be important in confounding effectiveness evaluation of the ES intervention will be descriptively reported and adjusted for during sensitivity analyses described in Section 11.9.1.1.

The number and proportion of women meeting inclusion in different analysis populations described in Section 10.1 will be reported, by treatment arm and overall. In addition, reasons for exclusions will be summarised. For instance, the following PP population exclusions will be considered:

a) Failed to meet any inclusion criterion as stipulated in the protocol but were included for some reason(s);
b) Switched randomised treatment in either direction for some reason(s);
c) Had a spontaneous pregnancy;
d) Embryo not generated for any other reason(s), such as because of failed fertilisation;
e) Cycle cancelled due for any other reason(s), such as because of insufficient follicle development;
f) Failed to use contraception prior to ES and their procedure could not be rescheduled;
g) Failed to receive treatment using fresh embryos;
h) Failed to receive IVF for some reason(s).
i) Were known to have received any protocols other than the antagonist or long (e.g., ultra-long)

11.2 The CONSORT flowchart: data completeness and disposition

Summarising data completeness is an integral part of good practice during trial reporting. Guided by the CONSORT statement for parallel-group individually randomised trials, the summary statistics in Table 2 will be calculated to construct a flowchart from screening, during follow-up, and to the analysis stage. The summaries will be made available to the trial monitoring committees during the conduct of the trial, presented by the centre and overall, and by treatment arm where appropriate. However, only the DMEC will have access to all summaries by treatment arm on request while the trial is ongoing within the remit of the agreed DMEC Charter.
It should be noted that all summaries, which may reveal the treatment effect such as pregnancy-related outcomes, will not be disclosed by treatment arm to the TMG and TSC members during the conduct of the trial.

Table 2: Information to construct a CONSORT flowchart

Event	Overall summary statistics to be reported
Screening	Number initially contacted and mode of contact: o Patient invitation letter, o Email invitation, o Patient information session, o Face-to-face at appointment, o Self-referral,

	○ Other. Number and proportion willing to participate (relative to those initially contacted) Number unwilling to take part with reasons: ○ Not interested, ○ Ineligible (group reasons if possible depending on observed data), ○ Illness, ○ Lack of time, ○ Unable to conduct visit or rearrange, ○ Involvement in competing study, ○ Unhappy to be randomised, ○ Prefers not to say, ○ Other.
Eligibility	Number and proportion eligible to take part (relative to those screened)
Ineligibility	Number excluded due to failure to meet inclusion criteria with reasons: ○ Previous trauma/surgery to the endometrium, ○ BMI ≥ 35 kg/m^2, ○ Known grade 4 endometriosis, ○ SET not expected, ○ Currently participating in any other fertility study involving medical/surgical intervention, ○ Other reasons.
Consent	Number and proportion consented (relative to those screened) Number not consented, but were deemed eligible with reasons ○ Not interested, ○ Illness, ○ Lack of time, ○ Involvement in competing study (E-Freeze/HABSelect/Other), ○ Not happy to be randomised, ○ Prefer not to say, ○ Other.
Randomisation	Overall number of women consented and randomised
	Numbers reported by treatment arm (where appropriate) Randomised to each intervention including the number who received and did not receive each intervention
ES procedure	Randomised to ES and received ES intervention Randomised to TAU, but received ES procedure for some reasons Randomised to ES, but did not receive it as per protocol with reasons: ○ Received ES elsewhere, ○ Failed to use contraception, ○ Achieved a spontaneous pregnancy, ○ Declined procedure, ○ Feeling unwell, ○ Other reasons.
Egg collection	Women whose eggs were collected Women whose eggs were uncollected with reasons (* end of study): ○ Empty follicles, ○ Early ovulation, ○ Other.
Fertilisation	Successful egg(s) fertilisation Unsuccessful egg(s) fertilisation (* end of study) Any embryos generated 1 day after fertilisation If no embryo(s) generated (* end of study)
Embryo transfer	Successful embryo transfer Unsuccessful embryo transfer with reasons (* end of study): ○ Abnormal uterine cavity,

	○ Hyperstimulation, ○ Unsuitable embryos to transfer, ○ Other.
8 weeks post egg collection: Pregnancy test	Had a pregnancy test Achieved biochemical pregnancy Not pregnant (* end of study)
Early pregnancy scan	Achieved biochemical pregnancy and positive foetal heartbeat Negative foetal heartbeat (* end of study) ○ Miscarriage, ○ Ectopic pregnancy, ○ Pregnancy of unknown location.
3 months post egg collection	Followed-up with ongoing pregnancy End of ongoing pregnancy with reasons (* end of study): ○ Miscarriage, ○ Stillbirth, ○ Pregnancy termination, ○ LTFU (* end of study), ○ Died (* end of study) ○ Withdrew consent (* end of study)
6 months post egg collection	Followed-up End of ongoing pregnancy with reasons (* end of study): ○ Stillbirth, ○ Pregnancy termination. LTFU (* end of study) Died (* end of study) Withdrew consent (* end of study)
10.5 months post egg collection	Followed-up (*end of study) Live birth Pre-term birth End of ongoing pregnancy with reasons (* end of study): ○ Stillbirth, ○ Pregnancy termination. LTFU (* end of study) Died (* end of study) Withdrew consent (* end of study)

Summaries which can reveal or enable the research team to guess the effect of the intervention effect such as spontaneous pregnancies, clinical pregnancy, miscarriages, stillbirths, ectopic pregnancy, and pregnancy of unknown location will not be presented to the TMG and TSC by treatment arm or overall during the course of the trial. Only pooled summaries will be presented where appropriate. The end of study can happen at any stage and will be reported with reasons: died; withdrew consent (not interested, illness, lack of time, unhappy with allocated treatment, prefer not to say, other), loss to follow-up (LTFU) and investigator decision; BMI: body mass index; SET: single embryo transfer; TAU: treatment as usual; ES: Endometrial Scratch.

The number of randomised women meeting the ITT criteria defined in Section 10.1 will be reported and presented in the CONSORT flowchart.

11.3 Data manipulation and definitions

The primary and secondary outcomes have been defined in Sections 4.2 and 4.3. Most of these outcomes are directly recorded on eCRFs so no additional data manipulations are required.

Age (years) will be calculated to one decimal place based on the date of oocyte removal and date of birth as given by:

$$\frac{(\text{Date of oocyte removal} - \text{DoB})}{365.25}$$

Gravidity relates to the number of times a woman has been pregnant. Parity is defined as the number of times a woman gave birth to a fetus with a gestational age of at least 24 weeks, regardless of whether the child was born alive or was stillborn. The "Medical History" eCRF contains the gestational age of the born fetus and the classification of the pregnancy outcome(s).

Preterm delivery is measured by live births after 24 weeks but before 37 weeks gestation. On the "pregnancy Outcome", gestation age and live birth outcomes are collected. A preterm delivery binary indicator will be created if pregnancy outcome is a live birth of gestation age ≥ 24 and < 37 weeks.

The delay (weeks) in IVF treatment after ES procedure will be computed as follows:

$$\text{Delay (weeks)} = \frac{(\text{FSH start date} - \text{ES procedure date})}{7}$$

The dates of the ES procedure and FSH are captured on eCRFs (ES procedure and Treatment cycle).

The duration of follow-up for participating mothers scaled to a year of follow-up (the exposure) will be calculated as follows for the analysis of safety outcomes:

$$\text{Woman's exposure} = \frac{(\text{discontinuation date} - \text{intervention date})}{365}$$

The above formula is applicable when reporting events that were experienced between uptake of the intervention (earlier date of ES or IVF if applicable) and study end. When reporting events that were recorded between the ES procedure and IVF, the calculation of woman's exposure (when necessary) will be based on the dates of the ES procedure and IVF.

Intervention date is the date when the ES procedure or IVF was received. FSH and ES date will be used as dates of intervention in the IVF and ES arm respectively. If a woman received both ES and IVF (most likely for the ES arm), the ES date will be used. Women who failed to receive any of the interventions (ES or TAU) will be excluded from safety analysis. If an unexpected AE or SAE date is greater than the study completion date then the AE date will be used as the last follow up (discontinuation date). Otherwise, the discontinuation date captured on the 'end of study involvement' form will be used as the last follow up date.

For PV bleeds, the AE and ES procedure dates will be used to ascertain whether the AE occurred within 2 days (~48 hrs) of the ES procedure.

11.4 Demographics and baseline characteristics of participating women

Summaries of the baseline variables relating to socio-demographics and characteristics of participating women captured on CRF will be reported by treatment arm and overall, depending on the distribution of variable under consideration, as shown in Table 3. Continuous variables will be summarised using minimum (min), maximum (max), mean and standard deviation (SD) or median, Interquartile Range (IQR), min and max depending on the skewness of the data. Categorical variables will be summarised using numbers and percentages in each category by treatment arm and overall. As for count variables, a decision on reporting approach will be made based on the underlying distribution of the pooled data. For instance, if the maximum number of counts is small, then a categorical variable will be derived and reported appropriately. Otherwise, the median (IQR) of the distribution of the count variable will be reported.

Table 3. Socio-demographic and characteristics of women at baseline by treatment arm

Variable	Scoring	TAU (n=xx)	ES (n=xx)	All (N=xx)
Site (fertility centre)	Sheffield	xx(xx%)	xx(xx%)	xx(xx%)
	Bradford	xx(xx%)	xx(xx%)	xx(xx%)
	Leicester	xx(xx%)	xx(xx%)	xx(xx%)
	Southampton	xx(xx%)	xx(xx%)	xx(xx%)
	Manchester	xx(xx%)	xx(xx%)	xx(xx%)
	Coventry and Warwick	xx(xx%)	xx(xx%)	xx(xx%)
	Birmingham	xx(xx%)	xx(xx%)	xx(xx%)
	Leeds	xx(xx%)	xx(xx%)	xx(xx%)
	Liverpool	xx(xx%)	xx(xx%)	xx(xx%)
	Homerton	xx(xx%)	xx(xx%)	xx(xx%)
	Newcastle	xx(xx%)	xx(xx%)	xx(xx%)
	Guys and Nottingham	xx(xx%)	xx(xx%)	xx(xx%)
	Oxford	xx(xx%)	xx(xx%)	xx(xx%)
	Wrightington	xx(xx%)	xx(xx%)	xx(xx%)
	Glasgow	xx(xx%)	xx(xx%)	xx(xx%)
	Gateshead	xx(xx%)	xx(xx%)	xx(xx%)
	South Tees	xx(xx%)	xx(xx%)	xx(xx%)
Age (years)	Mean (SD)	xx(xx)	xx(xx)	xx(xx)
	Median (IQR)	xx(xx to xx)	xx(xx to xx)	xx(xx to xx)
	Min to Max	xx to xx	xx to xx	xx to xx
Ethnicity [a]	White [b]	xx(xx%)	xx(xx%)	xx(xx%)
	Mixed/multiple ethnic groups [c]	xx(xx%)	xx(xx%)	xx(xx%)
	Asian/Asian British [d]	xx(xx%)	xx(xx%)	xx(xx%)
	Black/African/Caribbean/Black British [e]	xx(xx%)	xx(xx%)	xx(xx%)
	Other ethnic group [f]	xx(xx%)	xx(xx%)	xx(xx%)
	Prefer not to say	xx(xx%)	xx(xx%)	xx(xx%)
Current smoker	Yes	xx(xx%)	xx(xx%)	xx(xx%)
Number of cigarettes per week	≥1	xx(xx%)	xx(xx%)	xx(xx%)
	Mean (SD)	xx(xx)	xx(xx)	xx(xx)
	Median (IQR)	xx(xx to xx)	xx(xx)	xx(xx)
	Min to Max	xx to xx	xx to xx	xx to xx
Current recreational drug user	Yes	xx(xx%)	xx(xx%)	xx(xx%)

Alcohol intake (units per week)	≥1	xx(xx%)	xx(xx%)	xx(xx%)
	Mean (SD)	xx(xx)	xx(xx)	xx(xx)
	Median (IQR)	xx(xx to xx)	xx(xx)	xx(xx)
	Min to Max	xx to xx	xx to xx	xx to xx
BMI (kg/m^2)	Mean (SD)	xx(xx)	xx(xx)	xx(xx)
	Median (IQR)	xx(xx to xx)	xx(xx to xx)	xx(xx to xx)
	Min to Max	xx to xx	xx to xx	xx to xx
Planned method of fertilisation [j]	IVF	xx(xx%)	xx(xx%)	xx(xx%)
	ICSI	xx(xx%)	xx(xx%)	xx(xx%)
Planned treatment protocol [k]	Long	xx(xx%)	xx(xx%)	xx(xx%)
	Antagonist	xx(xx%)	xx(xx%)	xx(xx%)
Cycle programming	Yes	xx(xx%)	xx(xx%)	xx(xx%)
	Oral contraception (if yes only)	xx(xx%)	xx(xx%)	xx(xx%)
	Progestrogens (if yes only)	xx(xx%)	xx(xx%)	xx(xx%)
Duration of infertility (years) §	Mean(SD)	xx(xx)	xx(xx)	xx(xx)
	Median (IQR)	xx(xx to xx)	xx(xx to xx)	xx(xx to xx)
	Min to Max	xx to xx	xx to xx	xx to xx
History of fertility treatment	Yes	xx(xx%)	xx(xx%)	xx(xx%)
Fertility treatment received	IVF	xx(xx%)	xx(xx%)	xx(xx%)
	IUI	xx(xx%)	xx(xx%)	xx(xx%)
	Clomid	xx(xx%)	xx(xx%)	xx(xx%)
	Other	xx(xx%)	xx(xx%)	xx(xx%)
Number of previous pregnancies [g]	0	xx(xx%)	xx(xx%)	xx(xx%)
	1	xx(xx%)	xx(xx%)	xx(xx%)
	2	xx(xx%)	xx(xx%)	xx(xx%)
	3	xx(xx%)	xx(xx%)	xx(xx%)
	4	xx(xx%)	xx(xx%)	xx(xx%)
	≥5	xx(xx%)	xx(xx%)	xx(xx%)
Number of previous miscarriages [h]	0	xx(xx%)	xx(xx%)	xx(xx%)
	1	xx(xx%)	xx(xx%)	xx(xx%)
	2	xx(xx%)	xx(xx%)	xx(xx%)
	≥3	xx(xx%)	xx(xx%)	xx(xx%)
Parity [i]	0	xx(xx%)	xx(xx%)	xx(xx%)
	1	xx(xx%)	xx(xx%)	xx(xx%)
	2	xx(xx%)	xx(xx%)	xx(xx%)
	3	xx(xx%)	xx(xx%)	xx(xx%)
	4	xx(xx%)	xx(xx%)	xx(xx%)
	≥5	xx(xx%)	xx(xx%)	xx(xx%)

[a] The main ethnic groups could be collapsed depending on the observed distribution. [b] White: English/Welsh/Scottish/Northern Irish/British, Irish, Gypsy or Irish Traveller, and Any other White background; [c] Mixed/multiple ethnic groups: White and Black Caribbean, White and Black African, White and Asian, and Any other mixed/multiple ethnic groups background; [d] Asian/Asian British: Indian, Pakistani, Bangladeshi, Chinese, and Any other Asian background; [e] Black/African/Caribbean/Black British: African, Caribbean, and Any other Black/African/Caribbean/Black British background; [f] Other ethnic group: Arab, and Any other ethnic group; Prefer not to say. [g] Related to gravidity and only among women who experienced at least one previous pregnancy; [h] Only among women who experienced at least one miscarriage. [i] Parity defined by the number of times a woman gave birth to a fetus with a gestational age of at least 24 weeks, regardless of whether the child was born alive or was stillborn. § Different centres may have used different definitions in line with their routine practice. [g, i] Categories can be modified based on the distribution of the observed pooled baseline data. [h] Women who had more than 3 previous miscarriages are excluded. [j] Describe the number of women who changed the method of fertilisation from IVF to ICSI or vice versa. [k] Describe the number of women who changed the treatment protocol from antagonist to long or vice versa. IUI; Intrauterine insemination, IVF; In Vitro Fertilisation, ICSI; Intracytoplasmic Sperm Injection. Note that current smoker only relates to smoking cigarettes and not vaping.

11.5 Treatment cycle characteristics

In relation to a secondary objective (item 9) of Section 4.3, a detailed characterisation of women's treatment cycle will be reported at treatment, egg collection, fertilisation, and embryo transfer. Table 4 summarises details prior to egg collection.

Table 4. Characteristics of women's treatment prior to egg collection

Variable	Scoring	TAU	ES	All
		(n=xx)	(n=xx)	(N=xx)
Treatment Protocol	Antagonist	xx(xx%)	xx(xx%)	xx(xx%)
	Long (day 2)	xx(xx%)	xx(xx%)	xx(xx%)
	Long (day 21)	xx(xx%)	xx(xx%)	xx(xx%)
FSH drug used	Gonal F	xx(xx%)	xx(xx%)	xx(xx%)
	Merional	xx(xx%)	xx(xx%)	xx(xx%)
	Menopur	xx(xx%)	xx(xx%)	xx(xx%)
	Bemfola	xx(xx%)	xx(xx%)	xx(xx%)
	Other	xx(xx%)	xx(xx%)	xx(xx%)
Trigger	hCG	xx(xx%)	xx(xx%)	xx(xx%)
	Agonist	xx(xx%)	xx(xx%)	xx(xx%)
Number of days of FSH	1	xx(xx%)	xx(xx%)	xx(xx%)
	2	xx(xx%)	xx(xx%)	xx(xx%)
	…	…	…	…
Change in treatment protocol	Yes	xx(xx%)	xx(xx%)	xx(xx%)
Use of any other medications	Yes	xx(xx%)	xx(xx%)	xx(xx%)

hCG: human chorionic gonadotropin; FSH: follicle stimulating hormone. Note that FSH drugs could be used in combination so recorded drug combinations will be presented as separate categories.

Table 5 summarises details relating to treatment cycle at egg collection including the number of eggs collected and reasons for failure to collect eggs among some women.

Table 5. Characteristics of women's treatment cycle at egg collection

Variable	Scoring	TAU	ES	All
		(n=xx)	(n=xx)	(N=xx)
Eggs collected for fertilisation	Yes	xx(xx%)	xx(xx%)	xx(xx%)
Number of eggs collected (N=xx) [a]	1	xx(xx%)	xx(xx%)	xx(xx%)
	2	xx(xx%)	xx(xx%)	xx(xx%)
	…	….	….	….
Reasons for failure to collect eggs (N=xx) [b]	Empty follicles	xx(xx%)	xx(xx%)	xx(xx%)
	Early ovulation	xx(xx%)	xx(xx%)	xx(xx%)
	Other	xx(xx%)	xx(xx%)	xx(xx%)

[a] Denominator is the number of women whose eggs were successfully collected; [b] Denominator is the number of women whose eggs were not collected for some reason(s). The number of eggs collected could be presented as count data using mean (SD), median (IQR), minimum and maximum depending on the observed distribution.

Table 6 summarises the women's treatment cycle at fertilisation stratified by the allocated intervention arm.

Table 6. Characteristics of women's treatment cycle at fertilisation

Variable	Scoring	TAU (n=xx)	ES (n=xx)	All (N=xx)
Successful eggs fertilisation	Yes	xx(xx%)	xx(xx%)	xx(xx%)
Number of eggs fertilised [a] (n=xx)	1	xx(xx%)	xx(xx%)	xx(xx%)
	2	xx(xx%)	xx(xx%)	xx(xx%)
	…	…	…	….
Method of fertilisation	IVF	xx(xx%)	xx(xx%)	xx(xx%)
	ICSI	xx(xx%)	xx(xx%)	xx(xx%)
	Split ICSI	xx(xx%)	xx(xx%)	xx(xx%)
Change of fertilisation method	Yes	xx(xx%)	xx(xx%)	xx(xx%)
Reasons for change of fertilisation method [b] (n=xx)	Sperm quality	xx(xx%)	xx(xx%)	xx(xx%)
	Other	xx(xx%)	xx(xx%)	xx(xx%)
Embryos generated after fertilisation	Yes	xx(xx%)	xx(xx%)	xx(xx%)
Number of embryos generated [c] (n=xx)	1	xx(xx%)	xx(xx%)	xx(xx%)
	2	xx(xx%)	xx(xx%)	xx(xx%)
	…	….	…	…

[a] Denominator is the number of women whose eggs were fertilised; [b] Denominator is the number of women whose planned fertilisation method was changed for some reasons; [c] Denominator is the number of women whose embryos were generated after fertilisation.

Table 7 summarises the details of women's treatment cycle at embryo transfer stratified by the allocated treatment arm. This includes the success of embryo transfer, the number of embryos transferred, reasons for failure to transfer embryos, difficulties in embryo transfer, the day of embryo transfer, and quality of embryos as measured using the NEQAS and Gardners grading systems depending on routine clinical practices of recruiting sites.

Table 7. Characteristics of women's treatment cycle at embryo transfer

Variable	Scoring	TAU (n=xx)	ES (n=xx)	All (N=xx)
Embryo transferred	Yes	xx(xx%)	xx(xx%)	xx(xx%)
Reasons for failure to transfer embryo (N=xx) [a]	Abnormal uterine cavity	xx(xx%)	xx(xx%)	xx(xx%)
	Hyperstimulation	xx(xx%)	xx(xx%)	xx(xx%)
	No suitable embryos	xx(xx%)	xx(xx%)	xx(xx%)
	Other	xx(xx%)	xx(xx%)	xx(xx%)
Easy of embryo transfer [b]	Difficult	xx(xx%)	xx(xx%)	xx(xx%)
Number of embryos transferred [c]	Single	xx(xx%)	xx(xx%)	xx(xx%)
	Double	xx(xx%)	xx(xx%)	xx(xx%)
Day of embryo transfer [d]	2	xx(xx%)	xx(xx%)	xx(xx%)
	3	xx(xx%)	xx(xx%)	xx(xx%)
	4	xx(xx%)	xx(xx%)	xx(xx%)

	5	xx(xx%)	xx(xx%)	xx(xx%)
	6	xx(xx%)	xx(xx%)	xx(xx%)
Quality of embryos (blastocysts) transferred	Excellent	xx(xx%)	xx(xx%)	xx(xx%)
	Good	xx(xx%)	xx(xx%)	xx(xx%)
	Fair/freezable	xx(xx%)	xx(xx%)	xx(xx%)
	Fair	xx(xx%)	xx(xx%)	xx(xx%)
	Poor	xx(xx%)	xx(xx%)	xx(xx%)
Type of catheter used [f]	COOK	xx(xx%)	xx(xx%)	xx(xx%)
	Wallace plus obturator	xx(xx%)	xx(xx%)	xx(xx%)
	Wallace Sure-Pro	xx(xx%)	xx(xx%)	xx(xx%)
	Other	xx(xx%)	xx(xx%)	xx(xx%)
Blood on the tip of catheter [g]	Yes	xx(xx%)	xx(xx%)	xx(xx%)
Fluid in the endometrium [h]	Yes	xx(xx%)	xx(xx%)	xx(xx%)
Use of Volsellum [i]	Yes	xx(xx%)	xx(xx%)	xx(xx%)

[a] Denominator is the number of women whose embryos were not transferred; [b,c,d,e,f,g,h,i]; Denominator is the number of women whose embryos were transferred. [b] Only the difficult category is considered. [e] see appendix for bespoke embryo grading systems developed by an embryologist who was blinded to the trial results (Section 15).

11.6 Acceptability of ES procedure

One of the objectives of this study is to explore the acceptability of the ES procedure. Outcomes relating to the tolerability of the ES procedure (within 30 mins of the procedure) and perceived pain rating (within 30 mins of the procedure and 24 hrs and 7 days post-procedure) are recorded to address this objective. Tolerability is measured as a binary outcome (yes or no). Pain is measured on a rating scale of 0 (no pain) to 10 (worst pain imaginable). The proportion of women who tolerated the ES procedure with an exact 95% CI around this binomial proportion computed using the Wilson score method will be reported together as illustrated in Table 8 [19]. Two cases will be presented assuming: a) complete cases – only those with tolerability outcome data (CC), and b) those without tolerability outcome data did not tolerate the ES procedure (ITT worst-case).

Table 8. Acceptability of the ES procedure

Acceptability outcome	**Analysis set**	**Proportion of women (95% CI) (N=xx)**
Tolerability	CC	xx(xx%) [xx% to xx%]
	ITT worst-case	xx(xx%) [xx% to xx%]

CC: Complete case; ITT: Intention-to-treat; CI; Confidence Interval.

The distribution of pain rating scores at three timepoints following the ES procedure will be graphically displayed using a boxplot similar to the one shown in Figure 1. The distribution of the pain scores only at 24 hrs and 7 days will be stratified by eCRF version used (including unknown category when the version used was unclear). This is because there were some changes in the wording of the eCRF for clarification. A decision will be made to combine the pain scores depending on the observed distributions by eCRF version. For instance, if there are similarities in the

distribution of pain scores at different times between eCRF versions then pain score data will be combined and presented accordingly. Depending on the observed distribution of pain rating scores, data will be presented as means (SDs) or medians (IQRs). Here, only women who were randomised to the ES arm and received the procedure will be included for descriptive analysis.

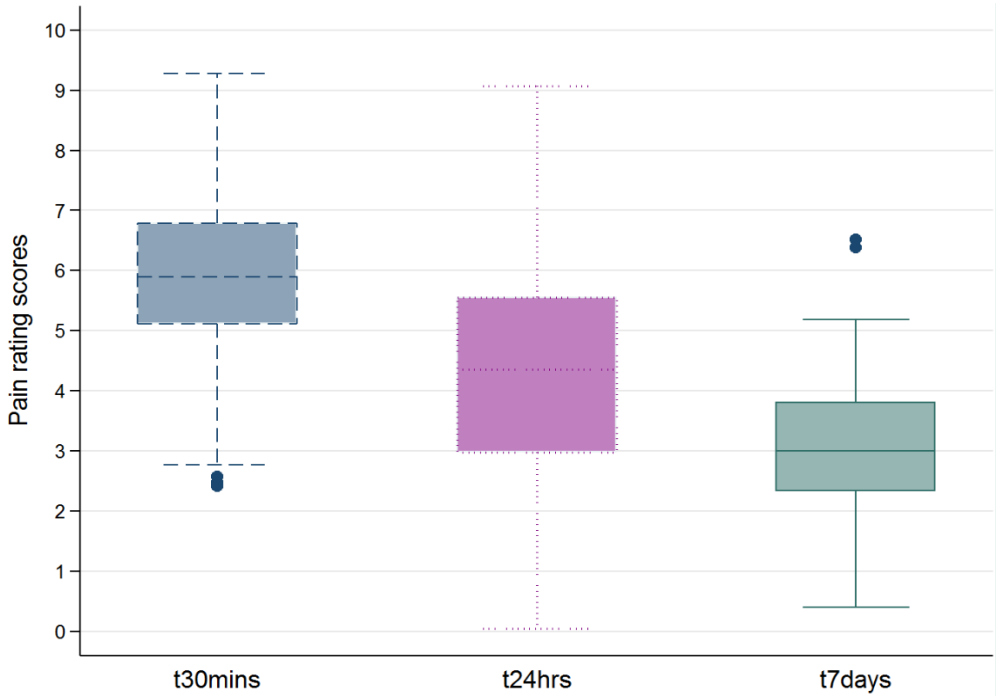

Figure 1. Dummy boxplot of pain rating scores at different timepoints following the ES procedure

Note that the higher the score the worse the pain (0=no pain and 10=the worst pain imaginable).

Noteworthy, the TMG noted that some women return ambiguous pain scores via text messages (e.g. with the comment "period pain, worse than usual"). Such responses are difficult to interpret whether the pain was purely due to period pain or exacerbated by the ES procedure. To facilitate informed interpretation about the acceptability of the ES procedure, the number and proportion of women with ambiguous pain scores at different timepoints will be reported and noted in the discussion of results. In addition, depending on the proportions of these women, at the discretion of the Trial Statistician, an additional descriptive analysis may be undertaken as described above but excluding women with ambiguous pain scores. Ambiguous scores may also be substituted by possible alternative responses recorded in the database as part of sensitivity analysis.

Some pain rating scores may be mistimed in the sense that the provided responses will not be closer to the expected timepoints. For example, although unlikely, some women may provide 7 days post-procedure data after 2 to 4 weeks of receiving a notification. The extent of recall bias on mistimed responses is unknown. Depending on the level of mistimed measurements, a sensitivity analysis will be performed by excluding mistimed responses – measurements

outside a certain 'acceptable' window. Members of the TMG blinded to other outcome data will aid decision-making about the 'acceptable' time window(s) of responses. A Trial Statistician will only provide data on the distributions of time responses and not the pain scores to the tasked TMG members to aid their decision-making process.

11.7 Distribution of the time to IVF following ES procedure

Women in the ES arm will ideally receive their IVF treatment approximately 1 week before the start of the IVF cycle when she is due to start stimulation. However, this may not be the case in some women due to other reasons. Although the effect of this delay in IVF treatment (start of the menstrual cycle) is unknown, it is expected that the effect of the ES procedure may diminish with increasing delay. It is therefore important to summarise and report the delay in IVF treatment following ES procedure. This delay (days or weeks) will be calculated as described in Section 11.3, presented graphically using either a boxplot or histogram and its distribution summarised accordingly. For instance, using min, median (IQR) and max if the distribution is skewed. Women who failed to receive scheduled IVF for some reason(s) will be excluded but their numbers noted.

11.8 Characteristics of completers and non-completers

The objective of this section is to explore the pattern of missing data and whether completers are systematically different from non-completers. Completers are defined as women whose primary outcome data relating to pregnancy and live birth is certainly known whereas non-completers are those with missing pregnancy and live birth data for some reason(s). Demographics and baseline characteristics of completers and non-completers will be descriptively explored. This exploratory analysis will be undertaken at the discretion of the Trial Statistician depending on the observed proportion of non-completers. Descriptive statistics of important potential prognostic factors of pregnancy or live birth outcomes will be presented stratified by the intervention arm and missing data status as illustrated in Table 9 and Table 10. The baseline variables include age, BMI, current smoking status, duration of infertility, recreational drug use, previous pregnancy, and previous miscarriages. Other variables presented in Table 3 may be considered. Furthermore, a univariable logistic regression model may be considered (at the discretion of the Trial Statistician) with the missing data status as the outcome (completers vs non-completers) and baseline covariates as explanatory variables.

Table 9. Continuous baseline characteristics by treatment arm and missing data status (completers vs non-completers)

Variable	Summary statistic	Completers			Non-completers		
		TAU (n=xx)	ES (n=xx)	All (n=xx)	TAU (n=xx)	ES (n=xx)	All (n=xx)
Age (years)	Mean(SD)	xx.x(xx.x)	xx.x(xx.x)	xx.x(xx.x)	xx.x(xx.x)	xx.x(xx.x)	xx.x(xx.x)
	Median(IQR)	xx.x(xx.x to xx.x)	xx.x(xx.x to xx.x)	xx.x(xx.x to xx.x)	xx.x(xx.x to xx.x)	xx.x(xx.x to xx.x)	xx.x(xx.x to xx.x)
BMI (kg/m^2)	Mean(SD)	xx.x(xx.x)	xx.x(xx.x)	xx.x(xx.x)	xx.x(xx.x)	xx.x(xx.x)	xx.x(xx.x)
	Median(IQR)	xx.x(xx.x to xx.x)	xx.x(xx.x to xx.x)	xx.x(xx.x to xx.x)	xx.x(xx.x to xx.x)	xx.x(xx.x to xx.x)	xx.x(xx.x to xx.x)
…	…	…	…	…	…	…	…

Table 10. Categorical baseline characteristics by treatment arm and missing data status (completers vs non-completers)

Variable	Scoring	Completers			Non-completers		
		TAU (n=xx)	ES (n=xx)	All (n=xx)	TAU (n=xx)	ES (n=xx)	All (n=xx)
Sex	Male	xx(xx.x%)	xx(xx.x%)	xx(xx.x%)	xx(xx.x%)	xx(xx.x%)	xx(xx.x%)
	Female	xx(xx.x%)	xx(xx.x%)	xx(xx.x%)	xx(xx.x%)	xx(xx.x%)	xx(xx.x%)
Current smoking status	Yes	xx(xx.x%)	xx(xx.x%)	xx(xx.x%)	xx(xx.x%)	xx(xx.x%)	xx(xx.x%)
	No	xx(xx.x%)	xx(xx.x%)	xx(xx.x%)	xx(xx.x%)	xx(xx.x%)	xx(xx.x%)
Recreational drug use	Yes	xx(xx.x%)	xx(xx.x%)	xx(xx.x%)	xx(xx.x%)	xx(xx.x%)	xx(xx.x%)
	No	xx(xx.x%)	xx(xx.x%)	xx(xx.x%)	xx(xx.x%)	xx(xx.x%)	xx(xx.x%)
…	…	…	…	…	..	…	…

11.9 Evaluation of the effectiveness of ES intervention

This section details statistical methods to be used to analyse the primary outcome and related sensitivity analysis in order to address the primary clinical effectiveness research question. Unless stated otherwise, the TAU will be used as the reference group for all statistical analyses.

11.9.1 Statistical Analysis of the primary endpoint

The primary analysis will be based on the ITT population as defined in Section 10.1. The unit of analysis is the woman randomised and not the IVF cycle. Under the strict ITT principle, all women will be included in the analysis once they are randomised as per the allocated treatment regardless of what happens after randomisation. For example, if a woman fails to receive IVF for some reason(s), they will be included in the ITT analysis and accounted for as described in Section 11.9.1.2.

The primary endpoint is the LBR defined by the number of live births after 24 weeks gestation within the 10.5-month post egg collection follow-up period relative to the number of women randomised. The number and proportions of live births relative to the number of women randomised will be reported by treatment arm and compared using a Chi-Square test. The effectiveness results will be reported as the difference in LBRs between arms with associated 95% CI calculated using Normal approximation to the Binomial distribution. Results will be presented as shown in Table 11. In consonance with the CONSORT guidance, the primary results will also be reported as MLE of the OR (Odds Ratio) with associated 95% CI based on a simple logistic regression model with the treatment arm as the only predictor. Furthermore, unadjusted MLE of the Relative Risk or Risk Ratio (RR) with associated 95% CI will be estimated and reported based on a log-binomial model – a generalised linear model (GLM) with a log link function and a binomial distribution [20,21].

Table 11. Presentation of the unadjusted primary analysis of the LBR

Outcome	TAU (n=xx)	ES (n=xx)	RD (95% CI)	RR (95% CI)	OR (95% CI)	p-value [a]
LBR	xx(xx%)	xx(xx%)	xx% (xx% to xx%)	xx% (xx% to xx%)	xx(xx to xx)	x.xxx

TAU as the reference group; [a] the same p-value using a Chi-Square test or a simple logistic regression model; OR=Odds Ratio; RD= Risk Difference (Difference in LBRs); RR=Relative Risk/Risk Ratio; CI=Confidence Interval; LBR=Live birth rate; TAU= 'treatment as usual'; ES=Endometrial Scratch.

11.9.1.1 Handling of stratification factors and multicentre trial data

Statistical literature recommends the adjustment for randomisation stratification factors during the analysis to increase precision [22]. Furthermore, in the event of notable between-group differences with respect to certain factors believed to confound the effectiveness evaluation of the ES intervention, the statistical analysis should account for such factors. With this in mind, a complementary adjusted analysis will be undertaken using a multiple

logistic regression model to account for stratification factors and potential confounding factors imbalanced at baseline. These variables will be treated as fixed factors in the multiple logistic regression model. An adjusted MLE of the OR with associated 95% CI and p-value will be reported, as displayed in Table 12, to support the primary results reported in Table 11. Any noted differences between the unadjusted and adjusted primary analyses will be highlighted. Fixed randomisation stratification factors to be included as additional predictors are:

a) Fertility Units (recruitment sites),
b) Treatment protocol (long or antagonist).

Based on clinical advice from the TMG, the following factors will be considered as additional covariates in the model for the supplementary adjusted analysis:

c) Age (years),
d) BMI (kg/m^2),
e) Smoking (yes/no),
f) Duration of infertility (years),
g) Previous pregnancy (yes/no).

Table 12. Presentation of the adjusted analysis for the primary endpoint: odds ratio scale

Outcome	**TAU**	**ES**	**Unadjusted OR (95% CI)** [a]	**p-value** [a]	**Adjusted OR (95% CI)** [b]	**p-value** [b]
	(n=xx)	(n=xx)				
LBR	xx(xx%)	xx(xx%)	xx(xx to xx)	x.xxx	xx(xx to xx)	x.xxx

[a] Results from Table 11; [b] Obtained from multiple logistic regression model; LBR: Live birth rate; OR=Odds Ratio; CI=Confidence Interval; TAU= 'treatment as usual'; ES=Endometrial Scratch.

In line with the CONSORT guidance, adjusted MLE of RR with associated 95% CI will be estimated using a log-binomial model – a generalised linear model (GLM) with a log link function and a binomial distribution [21,23]. This will be adjusted for randomisation stratification factors and covariates described above. The results will be reported as presented in Table 13.

Table 13. Presentation of the adjusted analysis for the primary endpoint: risk ratio scale

Outcome	**TAU**	**ES**	**Unadjusted RR (95% CI)** [a]	**p-value** [a]	**Adjusted RR (95% CI)** [b]	**p-value** [b]
	(n=xx)	(n=xx)				
LBR	xx(xx%)	xx(xx%)	xx(xx to xx)	x.xxx	xx(xx to xx)	x.xxx

[a] Results from Table 11; [b] Obtained from log-binomial regression model; LBR: Live birth rate; RR=Relative Risk/Risk Ratio; CI=Confidence Interval; TAU= 'treatment as usual'; ES=Endometrial Scratch.

Again in consonance with the CONSORT guidance, adjusted MLE of the risk difference (RD) will be estimated adjusted for randomisation stratification factors and covariates described above using one of the following approaches depending on convergence and model fitness;

i. a GLM either with a Binomial or Poisson distribution and log link function through estimation of margins [24],

ii. a GLM either with a Binomial or Poisson distribution and identity link function [25].

In either case, a GLM with a Binomial distribution and log link function will be the primary choice. In the case of a GLM with a Poisson distribution, robust adjusted standard errors will be used. Results will be reported as illustrated in Table 14.

Table 14. Presentation of the adjusted analysis for the primary endpoint: risk difference scale

Outcome	TAU (n=xx)	ES (n=xx)	Unadjusted RD (95% CI) [a]	p-value [a]	Adjusted RR (95% CI) [b]	p-value [b]
LBR	xx(xx%)	xx(xx%)	xx(xx to xx)	x.xxx	xx(xx to xx)	x.xxx

[a] Results from Table 11; [b] Obtained from a GLM as described above; LBR: Live birth rate; RD=Risk Difference; CI=Confidence Interval; TAU= 'treatment as usual'; ES=Endometrial Scratch.

11.9.1.2 Dealing with potential classification issues of the ITT population

During the ITT analysis, there are several potential classification issues for the primary outcome. These issues are and shall be dealt with as follows:

a) It is possible that some women will not undergo their IVF cycle for some reasons. These women will be included in the ITT analysis and assumed to have a treatment failure unless they have a known 'positive' pregnancy-related outcome. That is, they will contribute nothing to the numerator of the LBR;

b) A very small number of women may not receive the intended ES procedure as randomised prior to their IVF for some reasons. An attempt will be made to rearrange the ES procedure prior to starting IVF. If this attempt failed, women allocated to the ES arm but ended up receiving the IVF without the ES procedure will be analysed in the ES arm as part of the ITT population defined in Section 10.1. However, sensitivity analysis will be undertaken by excluding these women using the PP population, as described in Sections 10.1 and 11.9.2;

c) A small number of women may become spontaneously pregnant after randomisation prior to their IVF treatment and this may happen in both treatment arms. These women will be included in the primary ITT analysis as per randomised treatment arm and their outcome will contribute to the numerator of the LBR if the pregnancy achieved a live birth. However, sensitivity analysis will be undertaken by excluding these women using the PP population, as described in Sections 10.1 and 11.9.2;

d) Ideally, after 5 days, eggs fertilised in vitro develop into a blastocyst and a single blastocyst is transferred. Some women in both intervention arms may not have sufficient high-quality embryos to proceed to the blastocyst stage. These women will have eggs developed to embryos and after 2 to 3 days will have single or double embryo transfer rather than blastocyst transfer. These women will be

included in the ITT analysis as per randomised treatment. It is expected that the women included in this trial are most likely to be good candidates for single embryo transfer due to the age restriction inclusion criterion;

e) Women who fail to achieve pregnancy, had a miscarriage, or stillbirth will not be followed up after this point as governed by the trial protocol. Thus women who fail to get pregnant will contribute nothing to the numerator of pregnancy-related outcomes;
f) Women known to have died or LTFU prior to outcome assessment will be treated as treatment failures for related outcomes following death or LTFU.

11.9.1.3 Dealing with missing primary endpoint data

The presence of missing data relating to pregnancy and live births for some reasons pose problems and may introduce classification bias during the primary analysis. Related outcome(s) of trial dropouts prior to outcome(s) assessment cannot be known with certainty, unless for instance, if a woman is known to have died. A default conservative 'worst-case' scenario will be adopted as the primary approach for all the analyses unless stated otherwise. Here, a woman whose live-birth outcome is unknown for some reason(s) contribute to a negative outcome – a treatment failure. That is, they shall be assumed to have failed to produce a live birth. Supplementary sensitivity analysis will be performed as detailed in Section 11.9.2. A similar approach is adopted for secondary outcomes such as pregnancy and implantation.

11.9.1.4 Dealing with multiple births

Due to multiple pregnancies, there is a possibility that some pregnant women may achieve multiple births resulting in potential multiple live births outcomes from a single mother. For example, three multiple births from a single mother may yield 0, 1, 2 or 3 live births. Similarly, this may also happen to other multiple births related outcomes such as miscarriage or stillbirth. Based on clinical advice from the TMG, the following default approach will be adopted in dealing with multiple births for the primary analysis to create a binary outcome variable:

a) Assign one to the LBR numerator if there is at least one live birth from a single mother,
b) Assign zero to the numerator of the LBR if multiple births from a single mother resulted in no live birth(s).

That is, a live birth is counted as a single event regardless of how many babies are born during that live birth.

For sensitivity analysis, depending on the prevalence of multiple births from a single mother, a further complementary analysis will be undertaken accounting for multiple live births from a single mother by analysing the actual number of live births produced as repeated events rather than a binary outcome (1 if at least one live birth or 0 otherwise). This approach is detailed in Section 11.9.3.

11.9.2 Sensitivity analysis of the primary endpoint

In the case of missing birth outcome data, which is only likely to occur if the mother moves away, the following sensitivity approaches will be undertaken to supplement the default 'worst-case' scenario described in Section 11.9.1.3:

1. Imputation of the outcome data using a 'best-case' scenario, for instance, by assuming the woman within missing data had a successful live birth. Known deaths prior to outcome assessment will still be treated as treatment failures,
2. Analysis of available data based on the CC population by excluding those with missing outcome data.

The results based on the default 'worst-case' and 'best-case' scenarios will be reported and compared with the CC populations.

To supplement the primary analysis of LBR based on the ITT population, additional analysis will be undertaken using the PP population as defined in Section 10.1. A default 'worst-case' and 'best-case' scenarios will be used for this analysis as described in Sections 11.9.1.3 and 11.9.2, respectively. Furthermore, in the case of multiple births, a default approach described in Section 11.9.1.4 will be adopted. For this analysis, the proportions of the number of live births relative to the number of women randomised will be reported by treatment arm and compared using a Chi-Square test. The effectiveness results will be reported as the difference in the proportions of live births with associated 95% CI and p-value, as displayed in Table 11. As described in Section 11.9.1. In consonance with the CONSORT guidance, the primary results will also be reported as OR with associated 95% CI based on a simple logistic regression model with the treatment arm as the only predictor.

An adjusted analysis using a multiple logistic regression model will be undertaken as described in Section 11.9.1.1 to account for stratification factors and potential confounders depending on the observed imbalance between treatment arms. Table 15 summarises the primary and sensitivity analyses for the primary outcome (LBR) at 10.5 months from egg collection.

Table 15. Analysis sets for the primary outcome LBR at 10.5-months from egg collection

Analysis set	Description
Primary analysis	• ITT population, • Default 'worst-case' scenario for missing data as described in Section 11.9.1.3, • Multiple births dealt as described in Section 11.9.1.4 (items a and b).
Sensitivity analysis A	• ITT population, • 'Best-case' scenario for missing data as described in Section 11.9.2 (item 1), • Multiple births dealt with as described in Section 11.9.1.4 (items a and b).

Sensitivity analysis B	• CC population, • Multiple births dealt with as described in Section 11.9.1.4 (items a and b).
Sensitivity analysis C	• PP population, • 'Best-case' scenario for missing data as described in Section 11.9.2 (item 1), • Multiple births dealt with as described in Section 11.9.1.4 (items a and b).
Sensitivity analysis D	• PP population, • 'Worst-case' scenario for missing data as described in Section 11.9.1.3, • Multiple births dealt with as described in Section 11.9.1.4 (items a and b).
Sensitivity analysis E	• PP plus CC population, • Multiple births dealt with as described in Section 11.9.1.4 (items a and b).

Note that the 'best-case' scenario will not be used for women known to have died prior to outcome assessment. These will be assumed to have experienced treatment failure.

The results from the analysis sets summarised in Table 15 will be presented as shown in Table 12 to Table 14. In addition, forest plots of difference in proportions (RD), RR, and OR scales will be presented to aid visual interpretation, as illustrated in Figure 2 using RD scale.

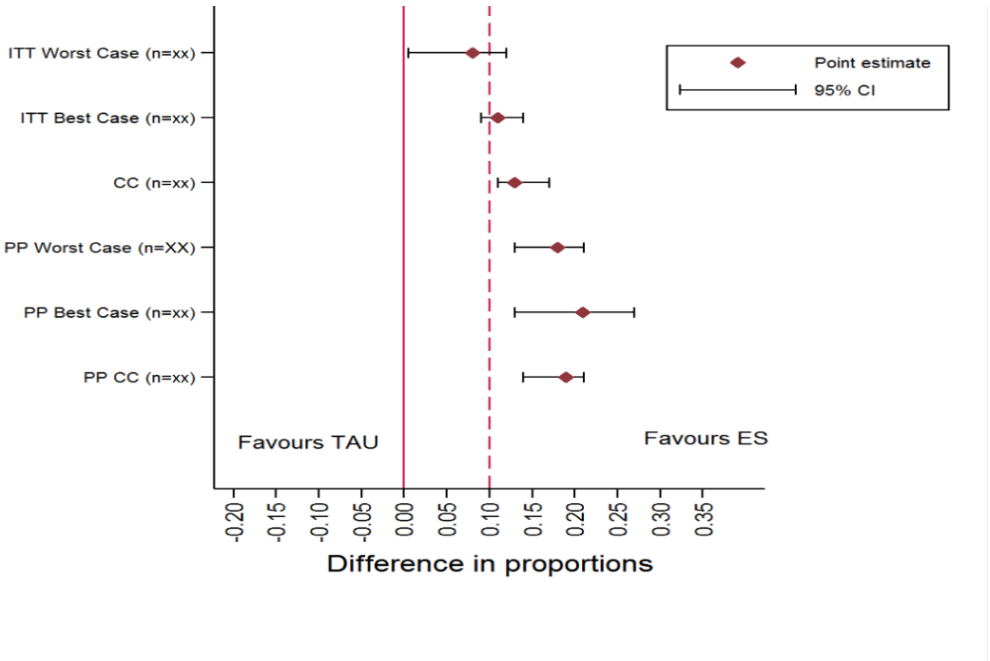

Figure 2. Dummy forest plot of sensitivity analysis on the primary endpoint (LBR): risk difference scale

11.9.3 Supplementary analysis to account for multiple births

So far, the analysis of LBR is based on a binary outcome as described in Section 11.9.1.4 ignoring multiple births from a single mother. This section describes additional analysis to account for potential multiple live births from a single

mother. Here, the goal is to investigate the effect of the ES intervention on the number of live births over a 10.5-month follow-up from egg collection (multiple live births rate).

The number of live births per single mother will be tabulated by treatment arm to explore its distribution. The number of live births per single mother will be modelled as counts using either a GLM with a log link function and:

1) Poisson distribution or,
2) Negative Binomial distribution in the presence of overdispersion.

The appropriateness of the Poisson model will be explored using descriptive statistics of the unconditional variance and mean. An appropriate model will be selected based on these descriptive findings. The analysis will only be for the ITT set using a conservative default 'worst-case' scenario for mothers whose live birth outcome is unknown for some reason(s), as described in Section 11.9.1.3. The mean incidence of live births per mother, Incidence Rate (IR), in each treatment arm over the study duration will be reported. The intervention effect will be reported as Incidence Rate Ratio (IRR) with associated 95% CI and p-value.

An adjusted analysis will be undertaken to account for stratification factors and potential confounders, described in Section 11.9.1.1, depending on the observed imbalance between treatment arms. The adjusted mean incidence of live births per mother, adjusted Incidence Rate (aIR), in each treatment arm over the study duration will be reported. The intervention effect will be reported as adjusted Incidence Rate Ratio (aIRR) with associated 95% CI and p-value.

11.10 Comparison of the number of eggs transferred and babies born

This section aims to explore the association between the number of eggs transferred and the number of babies born by a single mother. The outcome is the number of babies born by a single mother (count variable). The outcome will be summarised or plotted against the number of eggs transferred: a) overall and b) stratified by intervention arm. In addition, the outcome will be modelled in a multistage exploratory approach using a Poisson or Negative Binomial Regression model (as described in Section 11.9.3) as a function of:

1) Number of eggs transferred only,
2) Number of eggs transferred and intervention arm,
3) Number of eggs transferred, intervention arm, and interaction between the number of eggs transferred and intervention arm,

Results will be appropriately reported using forest plots of IRR (95% CI) stratified by the number of eggs transferred.

11.11 Exploratory subgroup evaluation of the ES intervention

The main objective of this section is to explore heterogeneity in the intervention effects in pre-specified subpopulations described in Section 10.2. Heterogeneity will be explored through an overall interaction test by fitting an interaction term between the intervention arm and subgroup indicator using a multiple logistic regression

model. The intervention effects (ORs and associated 95% CIs) will be obtained in each category of the subgroup, as shown in Table 16 and visually displayed using a forest plot [26] similar to Figure 2. The overall interaction test (intervention arm × subgroup) rather than calculating separate p-values within each category of the subgroup will be used to examine the strength of evidence for treatment heterogeneity across subgroups [27–29]. This analysis will be undertaken for the 'default' primary analysis of LBR using the ITT approach summarised in Table 15 in order to account for multiple births and missing data.

Table 16. Subgroup evaluation for the primary endpoint LBR using interaction tests

Variable	Subgroup	TAU	ES	OR (95% CI)	Overall Interaction test p-value
		n(%)	n(%)		
Fertilisation method [a]	IVF	xx(xx%)	xx(xx%)	xx (xx to xx)	
	ICSI	xx(xx%)	xx(xx%)	xx (xx to xx)	
	Split ICSI	xx(xx%)	xx(xx%)	xx (xx to xx)	x.xxx
Day of embryo transfer	2	xx(xx%)	xx(xx%)	xx (xx to xx)	
	3	xx(xx%)	xx(xx%)	xx (xx to xx)	
	4	xx(xx%)	xx(xx%)	xx (xx to xx)	
	5	xx(xx%)	xx(xx%)	xx (xx to xx)	
	6	xx(xx%)	xx(xx%)	xx (xx to xx)	x.xxx
Type of protocol	Long	xx(xx%)	xx(xx%)	xx (xx to xx)	
	Antagonist	xx(xx%)	xx(xx%)	xx (xx to xx)	x.xxx
Embryo transfer	Single	xx(xx%)	xx(xx%)	xx (xx to xx)	
	Double	xx(xx%)	xx(xx%)	xx (xx to xx)	x.xxx
Nature of embryo used	Fresh	xx(xx%)	xx(xx%)	xx (xx to xx)	
	Frozen	xx(xx%)	xx(xx%)	xx (xx to xx)	x.xxx
History of miscarriage	0-2	xx(xx%)	xx(xx%)	xx (xx to xx)	
	≥ 3	xx(xx%)	xx(xx%)	xx (xx to xx)	x.xxx
Cycle programming	No	xx(xx%)	xx(xx%)	xx (xx to xx)	
	Yes	xx(xx%)	xx(xx%)	xx (xx to xx)	x.xxx

[a] IVF and ICSI may be rarely used and if so, this subgroup of patients will be excluded.

The subgroup analysis will be undertaken for the primary outcome and important secondary outcomes presented in Section 4.3 (items 2 to 8) and highlighted in Section 11.13. The number of women known to have hydrosalpinx at any point during the trial will be reported.

11.12 Exploratory analysis of the effect IVF delay

Building on Section 11.7, the effect of the delay in IVF treatment following ES procedure (only in the ES arm) on pregnancy and live birth outcomes will be explored using a simple logistic regression. Delay (days or weeks) will be treated as the only continuous predictor. In addition, the log odds of achieving a positive outcome (pregnancy or live

birth) will be plotted against the delay in IVF treatment. Women who failed to receive IVF treatment for some reasons will be excluded since they would not have data on delay. Delay (days or weeks) could be transformed depending on model fit.

11.13 Effectiveness evaluation based on secondary endpoints

A number of secondary outcomes will be analysed to further examine the effect of the ES intervention. This will only be based on the ITT and PP using the default 'worse case' approach. These secondary outcomes as described in Section 4.3 are:

a) implantation rate,
b) ectopic pregnancy rate,
c) clinical pregnancy rate,
d) miscarriage rate,
e) multiple birth rate,
f) preterm delivery rate,
g) stillbirth rate.

These secondary outcomes will be treated as binary variables. The number of women and related event rate in each intervention arm will be estimated and reported. The difference in event rates between the intervention arms will be estimated and hypothesis test performed using a Chi-Square test. For consistency with the reporting of the primary outcome, a simple logistic regression model will also be fitted with the intervention arm as the only predictor. The results from a Chi-Square test (RD and associated 95% CIs), logistic regression model (OR and associated 95% CIs), and RR (95% CI) from a GLM described in Section 11.9.1 will be reported side by side, as displayed in Table 11. Here, a conservative default 'worse-case' scenario will be used to deal with missing outcome data as described for the primary analysis in Section 11.9.1.3. The presentation of the results based on the analysis of secondary endpoints will be reported as presented in Table 17. Note that exact regression methods (e.g. exact logistic regression model) may be considered at the discretion of the Trial Statistician if observed events are deemed rare.

Stillbirths, miscarriages, multiple births, and preterm delivery outcomes which relate to safety will be analysed repeatedly using the following denominators:

i. Number of women randomised similar to the approach adopted for pregnancy and live birth outcome,
ii. Number of pregnant women.

There is a possibility of multiple events of interest on some outcomes that are influenced by multiple pregnancies or multiple births from a single mother, such as the number of stillbirths. By default, binary variables will be created as follows:

a) Assign one to the numerator if there is at least one outcome of interest (such as stillbirth or miscarriage) from a single mother,
b) Assign zero to the numerator if the outcome a single mother resulted in no outcome of interest.

Depending on the observed distribution of repeated events from outcomes influenced by multiple pregnancies, additional sensitivity analysis will be undertaken using the approach described in Section 11.9.3. This approach will not be considered if the observed frequency of repeated events is subjectively viewed as negligible to alter the interpretation of results based on the default approach.

Table 17. Unadjusted effectiveness of ES intervention based on secondary outcomes

Secondary outcome	TAU (n=xx)	ES (n=xx)	RD (95% CI) [a]	RR (95% CI) [b]	Unadjusted OR (95% CI) [c]	p-value [d]
Implantation rate	xx(xx%)	xx(xx%)	xx% (xx% to xx%)	xx(xx to xx)	xx(xx to xx)	x.xxx
Ectopic pregnancy rate	xx(xx%)	xx(xx%)	xx% (xx% to xx%)	xx(xx to xx)	xx(xx to xx)	x.xxx
Clinical pregnancy rate	xx(xx%)	xx(xx%)	xx% (xx% to xx%)	xx(xx to xx)	xx(xx to xx)	x.xxx
Miscarriage rate [e]	xx(xx%)	xx(xx%)	xx% (xx% to xx%)	xx(xx to xx)	xx(xx to xx)	x.xxx
Multiple birth rate [e]	xx(xx%)	xx(xx%)	xx% (xx% to xx%)	xx(xx to xx)	xx(xx to xx)	x.xxx
Preterm delivery rate [e]	xx(xx%)	xx(xx%)	xx% (xx% to xx%)	xx(xx to xx)	xx(xx to xx)	x.xxx
Stillbirth rate [e]	xx(xx%)	xx(xx%)	xx% (xx% to xx%)	xx(xx to xx)	xx(xx to xx)	x.xxx

[a] Results from Chi-Square test; [b] Results from a GLM as described in Section 11.9.1; [c] Results from simple logistic regression model; [d] Results from Chi-Square or simple logistic regression model; [e] Repeated using the number of pregnant women as the denominator.

Adjusted analysis to account for the randomisation stratification factors and other important baseline covariance will be undertaken as described for the primary endpoint in Section 11.9.1.1. Results will be reported in RR, RD, and OR scales as presented in Table 18 to Table 20.

Table 18. Unadjusted effectiveness of ES intervention based on secondary outcomes: odds ratio scale

Secondary outcome	TAU (n=xx)	ES (n=xx)	Unadjusted OR (95% CI) [a]	Unadjusted p-value [a]	Adjusted OR (95% CI) [b]	Adjusted p-value [b]
Implantation rate	xx(xx%)	xx(xx%)	xx(xx to xx)	x.xxx	xx(xx to xx)	x.xxx
Ectopic pregnancy rate	xx(xx%)	xx(xx%)	xx(xx to xx)	x.xxx	xx(xx to xx)	x.xxx

Clinical pregnancy rate	xx(xx%)	xx(xx%)	xx(xx to xx)	x.xxx	xx(xx to xx)	x.xxx
Miscarriage rate [c]	xx(xx%)	xx(xx%)	xx(xx to xx)	x.xxx	xx(xx to xx)	x.xxx
Multiple birth rate [c]	xx(xx%)	xx(xx%)	xx(xx to xx)	x.xxx	xx(xx to xx)	x.xxx
Preterm delivery rate [c]	xx(xx%)	xx(xx%)	xx(xx to xx)	x.xxx	xx(xx to xx)	x.xxx
Stillbirth rate [c]	xx(xx%)	xx(xx%)	xx(xx to xx)	x.xxx	xx(xx to xx)	x.xxx

[a] Results from Chi-Square test presented in Table 17; [b] Adjusted results from multiple logistic regression model; [c] Repeated using the number of pregnant women as the denominator.

Table 19. Unadjusted effectiveness of ES intervention based on secondary outcomes: relative risk scale

Secondary outcome	**TAU (n=xx)**	**ES (n=xx)**	**Unadjusted RR (95% CI) [a]**	**Unadjusted p-value [a]**	**Adjusted RR (95% CI) [a]**	**Adjusted p-value [b]**
Implantation rate	xx(xx%)	xx(xx%)	xx (xx to xx)	x.xxx	xx (xx to xx)	x.xxx
Ectopic pregnancy rate	xx(xx%)	xx(xx%)	xx (xx to xx)	x.xxx	xx (xx to xx)	x.xxx
Clinical pregnancy rate	xx(xx%)	xx(xx%)	xx (xx to xx)	x.xxx	xx (xx to xx)	x.xxx
Miscarriage rate [c]	xx(xx%)	xx(xx%)	xx (xx to xx)	x.xxx	xx (xx to xx)	x.xxx
Multiple birth rate [c]	xx(xx%)	xx(xx%)	xx (xx to xx)	x.xxx	xx (xx to xx)	x.xxx
Preterm delivery rate [c]	xx(xx%)	xx(xx%)	xx (xx to xx)	x.xxx	xx (xx to xx)	x.xxx
Still birth rate [c]	xx(xx%)	xx(xx%)	xx (xx to xx)	x.xxx	xx (xx to xx)	x.xxx

[a] Unadjusted results presented in Table 17; [b] Adjusted results from a GLM as described in Section 11.9.1; [c] Repeated using the number of pregnant women as the denominator.

Table 20. Unadjusted effectiveness of ES intervention based on secondary outcomes: risk difference scale

Secondary outcome	**TAU (n=xx)**	**ES (n=xx)**	**Unadjusted RD (95% CI) [a]**	**Unadjusted p-value [b]**	**Adjusted RD (95% CI) [a]**	**Adjusted p-value [b]**
Implantation rate	xx(xx%)	xx(xx%)	xx% (xx% to xx%)	x.xxx	xx% (xx% to xx%)	x.xxx
Ectopic pregnancy rate	xx(xx%)	xx(xx%)	xx% (xx% to xx%)	x.xxx	xx% (xx% to xx%)	x.xxx
Clinical pregnancy rate	xx(xx%)	xx(xx%)	xx% (xx% to xx%)	x.xxx	xx% (xx% to xx%)	x.xxx
Miscarriage rate [c]	xx(xx%)	xx(xx%)	xx% (xx% to xx%)	x.xxx	xx% (xx% to xx%)	x.xxx
Multiple birth rate [c]	xx(xx%)	xx(xx%)	xx% (xx% to xx%)	x.xxx	xx% (xx% to xx%)	x.xxx
Preterm delivery rate [c]	xx(xx%)	xx(xx%)	xx% (xx% to xx%)	x.xxx	xx% (xx% to xx%)	x.xxx
Stillbirth rate [c]	xx(xx%)	xx(xx%)	xx% (xx% to xx%)	x.xxx	xx% (xx% to xx%)	x.xxx

[a] Unadjusted results presented in Table 17; [b] Adjusted results from a GLM as described in Section 11.9.1; [c] Repeated using the number of pregnant women as the denominator.

11.14 Analysis of safety outcomes

Safety outcomes are recorded as AEs and SAEs for participating women and born foetuses. It should be noted that participating women are not followed-up for the same duration as described in the protocol. As a result, the total number of unexpected AEs and SAEs depends on the exposure (Section 11.3), which is the duration of follow-up for that particular woman, as highlighted in Table 2. For instance, some women will only be followed-up to the point of a negative outcome such as a miscarriage or stillbirth instead of the entire study duration due to ethical considerations. It is also possible that some women may experience multiple or repeated unexpected AEs or SAEs during follow-up. Thus, the reporting of only the total number of unexpected AEs or SAEs between treatment arms may give a misleading picture of the safety profile of the ES procedure compared to the TAU arm. In addition, due to the delay between randomisation and the start of the ES or IVF, the research team is primarily interested in unexpected AEs and SAEs that occur after ES procedure or IVF depending on the intervention arm women are allocated and/or received. As a result, safety analysis will focus on safety events that occur between the delivery of the:

1. interventions and study end,
2. ES procedure and IVF only in the ES arm.

That is, women who did not receive any of the interventions will be excluded to minimise overreporting of events that are known to be completely unrelated to the interventions. The second item is only relevant among women who received both the ES procedure and IVF procedure so no comparison will be made.

In summary, in addition to the total number of unexpected AEs and SAEs recorded, the analysis and reporting of unexpected AEs and SAEs will account for the exposure (duration of follow-up) and repeated events per participating woman where appropriate. For expected AEs, repeated events are not recorded in the database so only the total number and proportions will be reported. For item 2 above, only the total number of events may be reported at the discretion of the Trial Statistician after looking at the data depending on the prevalence of repeated events and variability of the distribution of time between the ES procedure and IVF.

All safety analysis and sensitivity analysis will be based on safety analysis populations described in Section 10.

11.14.1 Description of AEs and SAEs for participating women

Descriptive summaries of AEs and SAEs will be reported by treatment arm and overall without formal statistical hypothesis tests. AEs and SAEs will be reported based on the actual intervention the women received. For example, if a woman is randomised to ES but fails to receive it and went on to receive IVF then they will be treated as received

TAU. Women who fail to receive the ES and IVF when allocated to ES or IVF when allocated to IVF will be excluded as long as they do not receive any of the interventions.

Whenever repeated events are reported, the total number of events will be reported with incidences. For non-repeated events, numbers and percentages will be reported. Repeated unexpected AEs or SAEs will be analysed using a Poisson model or Negative Binomial Model in the presence of overdispersion depending on the observed distribution of repeated events. This will account for the women exposure as described under Section 11.3). Results will be reported as the total number of repeated events, IR (average events per woman per year), and IRR and associated 95% CI where appropriate, as illustrated in Table 21 and Table 22.

For recorded AEs, the following summaries will be reported and presented as illustrated in Table 21:

a) Total number of all AEs recorded (expected and unexpected) and incidence rate by treatment arm and overall,
b) Number and proportion of women who recorded at least one AEs (expected and unexpected) by treatment arm and overall,
c) Total number of unexpected AEs recorded and the incidence by treatment arm and overall,
d) Number and proportion of women who recorded at least one unexpected AEs by treatment arm and overall,
e) Total number of expected AEs recorded and the incidence by treatment arm and overall,
f) Number and proportion of women who recorded at least one expected AEs by treatment arm and overall,
g) Total number of expected AEs by type (category as agreed by the TMG) with an incidence rate.

Table 21. Summary of AEs experienced by women and their description

Variable	**Classification**	**Treatment received**				
		TAU		**ES**		
		(n=xx)	**IR**	**(n=xx)**	**IR**	**IRR (95% CI)**
Number of all AEs	Including repeated events	Xx	xx	xx	xx	xx (xx to xx)
Women with ≥1 AE	Any AE	xx(xx%)		xx(xx%)		
Number of all expected AEs	Including repeated events	Xx	xx	xx	xx	xx (xx to xx)
Women with ≥1 AE expected AE	Any expected AE	xx(xx%)		xx(xx%)		
			[n, IR]		[n, IR]	
Type of expected AE	Abdominal pain	xx(xx%)		xx(xx%)		
	Clicky hip	xx(xx%)		xx(xx%)		
	Conjunctivitis	xx(xx%)		xx(xx%)		
	Constipation			…		
	Cough	xx(xx%)		xx(xx%)		
	Diarrhoea	xx(xx%)		xx(xx%)		
	Dizziness	xx(xx%)		xx(xx%)		
	Facial pain	xx(xx%)		xx(xx%)		
	Gestational diabetes	xx(xx%)		xx(xx%)		

	…	…		….		
	…	…		….		
	…	…		….		
	Vaginal infection	xx(xx%)		xx(xx%)		
	Viral infection	xx(xx%)		xx(xx%)		
Number of all unexpected AEs	Including repeated events	xx	xx	xx	xx	xx (xx to xx)
Women with ≥1 unexpected AE	Any unexpected AE	xx(xx%)		xx(xx%)		

[n, IR] = [total number of repeated events, the average incidence rate per woman per year]; note that repeated expected events are not recorded. For per vaginal (PV) bleeding, total events regardless of when they happened will be reported. In addition, PV bleeds that occurred within 48 hrs (~2days) after the ES procedure will be reported as this is clinically more important.

As for recorded SAEs, the following summaries will be reported and presented as illustrated in Table 22:

a) Total number of all recorded SAEs (expected and unexpected) and incidence rate per treatment arm and overall,
b) Number and proportion of women who recorded at one SAE by treatment arm and overall,
c) Total number of all SAE with incidence rate stratified by seriousness (death, life-threatening, inpatient hospitalisation, prolonged hospitalisation, and persistent or significant disability or incapacity) ,
d) Total number of all SAE with incidence rate as defined by frequency (isolated, intermittent, continuous, and unknown),
e) Total number of all SAE with incidence rate as defined by intensity (mild, moderate, and severe) ,
f) Number and proportion of women who recorded at least one SAE as defined by outcome (recovered, improved, unchanged, deteriorated, persisted, and death). The total number of SAEs by outcome with incidence rate may be considered depending on the frequency of events,
g) Total number of expected recorded SAEs and incidence rate per treatment arm and overall,
h) Number and proportion of women who recorded at one expected SAE by treatment arm and overall,
i) Total number of SAEs related to the ES procedure and incidence,
j) Number of SAEs related to SAE procedure and incidence rate as defined by relationship (definite, probable, possible, unlikely, unrelated, and not assessable).

Table 22. Summary of SAEs experienced by women and their description

Variable	**Classification**	**Treatment received**				
		TAU		**ES**		
		(n=xx)	**IR**	**(n=xx)**	**IR**	**IRR (95% CI)**
Number of all SAEs	Including repeated events	Xx	xx	Xx	xx	xx (xx to xx)
Women with ≥1 SAE	Any SAE	xx(xx%)		xx(xx%)		
			[n, IR]		[n, IR]	
Serious	Death	xx(xx%)	[xx, xx]	xx(xx%)	[xx, xx]	
	Life threatening	xx(xx%)	[xx, xx]	xx(xx%)	[xx, xx]	
	Inpatient hospitalisation	xx(xx%)	[xx, xx]	xx(xx%)	[xx, xx]	
	Prolonged hospitalisation	xx(xx%)	[xx, xx]	xx(xx%)	[xx, xx]	
	Persistent or significant disability/incapacity	xx(xx%)	[xx, xx]	xx(xx%)	[xx, xx]	
Frequency	Isolated	xx(xx%)	[xx, xx]	xx(xx%)	[xx, xx]	

		Intermittent	xx(xx%)	[xx, xx]	xx(xx%)	[xx, xx]	
		Continuous	xx(xx%)	[xx, xx]	xx(xx%)	[xx, xx]	
		Unknown	xx(xx%)	[xx, xx]	xx(xx%)	[xx, xx]	
	Intensity	Mild	xx(xx%)	[xx, xx]	xx(xx%)	[xx, xx]	
		Moderate	xx(xx%)	[xx, xx]	xx(xx%)	[xx, xx]	
		Severe	xx(xx%)	[xx, xx]	xx(xx%)	[xx, xx]	
	Outcome	Recovered	xx(xx%)	[xx, xx]	xx(xx%)	[xx, xx]	
		Improved	xx(xx%)	[xx, xx]	xx(xx%)	[xx, xx]	
		Unchanged	xx(xx%)	[xx, xx]	xx(xx%)	[xx, xx]	
		Deterioration	xx(xx%)	[xx, xx]	xx(xx%)	[xx, xx]	
		Persisted	xx(xx%)	[xx, xx]	xx(xx%)	[xx, xx]	
		Death	xx(xx%)	[xx, xx]	xx(xx%)	[xx, xx]	
All expected SAEs		Including repeated events	Xx	xx	Xx	xx	xx (xx to xx)
Women with ≥1 expected SAEs		Any expected SAEs	xx(xx%)		xx(xx%)		
SAEs relationship to ES		Definite			xx(xx%)	[xx, xx]	
		Probable			xx(xx%)	[xx, xx]	
		Possible			xx(xx%)	[xx, xx]	
		Unlikely			xx(xx%)	[xx, xx]	
		Unrelated			xx(xx%)	[xx, xx]	
		Not assessed			xx(xx%)	[xx, xx]	

[n, IR] = [total number of repeated events, the average incidence rate per woman per year]

A detailed listing of AEs and SAEs will be made available to the DMEC on their request at any point during the trial. In addition, the listings will be provided as part of clinical trials registry reporting at the end of the trial. This will be done at the level of the mothers and born babies. The listing will include:

a) Allocated treatment arm and received treatment,
b) Description of the event,
c) Category,
d) Seriousness,
e) Frequency,
f) Intensity,
g) Relationship to the ES procedure,
h) Action was taken and outcome.

11.14.2 Description of SAEs for born foetuses

In this trial, neonatal AEs are not recorded. Two types of neonatal SAEs that are recorded; severe congenital abnormality and neonatal death including the date of diagnosis and death, respectively. These SAEs will be reported based on the number of pregnant women as the denominator in each treatment arm. The number of babies (repeated events) who experienced each SAE classification will be summarised by treatment arm. We will also report the proportion of babies per mother who experienced at least one SAE classification by treatment arm. For each SAEs category, the IR and IRR with 95% CI will be estimated using a Poisson or Negative Binomial distribution

depending on the underlying distribution of repeated SAEs. This analysis will be performed on pregnant women only. The differences in proportions of babies per mother who experienced at least one SAE category will be estimated and related CIs computed using Normal approximation to the Binomial distribution.

Table 23. Summary of SAEs in born foetuses and their description

Variable	SAE classification	TAU (N=xx)	IR	ES (N=xx)	IR	IRR (95% CI)
Total number of babies who suffered:	Severe congenital abnormality	Xx	xx	Xx	xx	xx(xx to xx)
	Neonatal death	Xx	xx	Xx	xx	xx(xx to xx)
		n(%)		n(%)		Differences in proportions (95% CI)
Babies with ≥ 1 SAE per women	Severe congenital abnormality	xx(xx%)		xx(xx%)		xx(xx% to xx%)
	Neonatal death	xx(xx%)		xx(xx%)		xx(xx% to xx%)
Unexpected SAEs						
Babies with ≥ 1 SAE per women	Low birth weight	xx(xx%)		xx(xx%)		xx(xx% to xx%)
	Very low birth weight	xx(xx%)		xx(xx%)		xx(xx% to xx%)
	Large for gestational age	xx(xx%)		xx(xx%)		xx(xx% to xx%)
	Preterm delivery	xx(xx%)		xx(xx%)		xx(xx% to xx%)
	Very preterm delivery	xx(xx%)		xx(xx%)		xx(xx% to xx%)
	Small for gestational age	xx(xx%)		xx(xx%)		xx(xx% to xx%)

Low birth weight defined as weighing 2499g or less at birth or <10th centile; very low birth weight defined by weight less than 1500g at birth or <5th centile; large for gestational age defined as estimated fetal weight above the 95th centile for gestation (at birth); preterm delivery defined as delivery between 24 weeks and 37 weeks; very preterm delivery defined as delivery before 24 completed weeks and; small for gestational age defined as estimated weight less than the 10th centile.

To obtain centiles, anthropometric measurements are converted using the World Health Organization (WHO) standards [30]. That is, for a given gestation age (in days), weight (in kilograms), and sex (male or female), corresponding centiles are estimated using growth curves and these are then used to classify related SAEs of born babies stated above (such as low birth weight, small for gestational age, etc). An R package (hbgd) will be used to convert anthropometric measurements to centiles as described here: https://hbgdki.github.io/hbgd/#growth-standards.

12. Statistical Model Selection and Diagnostics

The Hosmer-Lemeshow goodness-of-fit test will be used to investigate goodness of fit of the fitted logistic regression model. Residuals such as deviance, pearson, and standardised will be used to identify potential outliers and influential observations. The ratio of the mean and variance of the outcome data under consideration with the aid of graphical plots (such as histograms) will be used to investigate overdispersion. If the results suggest the existence of overdispersion which violates the assumption of the Poisson Regression model, then a Negative Binomial Regression model will be used to model repeated count outcome data to account for overdispersion [31]. For instance, if the ratio of the mean to the variance is greater than 1.5.

13. Implementation of the SAP

This SAP will be used as a work description of the Trial Statistician in consultation with the Senior Trial Statistician. There will be no analysis to be undertaken until after the sign-off of this SAP by relevant personnel. Data will be released by the data management after sign-off to the Trial Statistician (after data freeze), given a window period to query any spurious data and initiate data lock before actual analysis. At this point, no changes will be allowed on the database. Unblinded DMEC reports will be produced by the Sheffield CTRU Statistician or data management team.

14. Appendix

Embryo quality grade	Embryo grading system		
	Gardners	**New NEQAS**	**Old NEQAS**
Excellent	A/A	A/A	5/3
Very good	A/B, B/A	A/B, B/A	5/2, 4/3, 3/4, 4/4
Good	B/B	B/B	4/2
Fair + freezable	A/C , C/A	BC/CB, A/C	3/3, 3/2
Fair	B/C, C/B	C/C, A/D	3/1, 4/1, 5/1 2/3, 2/2, 1/3
Poor ǂ	C/C, Degree of expansion 2 and X/X	C/D, D/C, D/D, Degree of expansion 2 and X/X	2/1, 1/1, 1/2
Early blastocyst ǂ	No grades provided (X/X/X) Degree of expansion 1 and any other TE or ICM grade	No grades provided (X/X/X) Degree of expansion 1 and any other TE or ICM grade	No grades provided (X/X/X) Degree of expansion 1 and any other TE or ICM grade

ǂ X means grade not provided; ICM, inner cell mass; TE, trophectoderm; Order of presentation: Y/Z means inner cell mass of Y and trophectoderm score of Z ignoring the degree of expansion score

Embryo quality grade	Cleavage embryo grading
Excellent	Day 2: 4/4/4, 3/4/4. Day 3: 8/4/4.
Good	Day 2: 5/4/4, 5/3/4, 5/4/3, 5/3/3, 4/3/4, 4/4/3, 4/3/3. Day 3: 10/4/4, 10/4/3, 10/3/4, 10/3/3, 9/4/4, 9/4/3, 9/3/4, 9/3/3, 8/4/3, 8/3/4, 8/3/3, 7/4/4, 7/4/3, 7/3/3, 6/4/4, 6/3/4, 6/4/3, 6/3/3, 7/3/4.
Fair	Day 2: 5/2/3, 5/2/4, 4/2/3, 3/3/4, 3/3/3, 3/2/3, 4/2/4. Day 3: 6/2/4, 8/4/2.
Poor quality	Day 2: All ≥6 cell number combinations, 5/3/2, 5/2/2, 4/3/2, 4/2/2, 3/4/3, 3/3/2, 3/2/2. Day 3: All ≥11 cell number combinations, 10/3/2, 10/2/3, 102/2, 9/3/2, 9/2/3, 9/2/2, 8/3/2, 8/2/3, 8/2/2, 7/3/2, 7/2/3, 7/2/2, 6/3/2, 6/2/30, 6/2/2.
Very poor quality	Day 2: 5/2/1, 5/1/2, 5/1/1, 4/2/1, 4/1/2, 4/1/1, 3/2/1, 3/1/2, 3/1/1, 3/1/3. Day 3: All -/1/1 combinations, 5/2/1.
Slow	Day 2: All 2 cell number combinations. Day 3: All ≤ 5 cell combinations except -/1/1, 5/4/4, 5/3/4, 5/4/3, 5/3/3, 5/2/3, 5/3/2, 5/2/2, 5/2/3
Arrested development	Graded as cleavage at Day 5 of development

Day 2, 3 and 5 are the days of embryo transfer; for the old NEQAS, the order of representation means cell number/shape score/fragmentation score; for the new NEQAS, the order of representation means: cell number/blastomere size/fragmentation score. "-/Y/Z" means any cell number combination with blastomere or shape size of Y and fragmentation score of Z.

15. Summary of changes from the previous version

Noteworthy changes to version 1 of the SAP are summarised in Table 24.

Table 24. Summary of key amendments made to version 1 of the SAP

Version	Date approved	Modifications (with sections)	When
1.0	24/02/2017	Not applicable	Not applicable
2.0	13/11/2019		
2.0		Throughout the SAP, 5 months after birth has been renamed "6 weeks post-partum" in line with a protocol amendment.	Not relevant.
2.0		Section 4.3 now includes the Gardner embryo grading system that was also used by some centres.	Prior to unblinded and blinded review
2.0		Section 6 now includes eligibility screening for completeness.	Not relevant
2.0		Section 7.1 clarifies hat block sizes used during randomisation will be disclosed after trial completion during reporting.	Prior to unblinded and blinded review
2.0		Section 9.1; for completeness as this very unlikely to happen, we clarified the ITT set by stating that women who withdrew consent and explicitly stated that their data should not be used will be excluded.	Prior to unblinded and blinded review
		Section 11.1 updated safety analysis set to reflect that women who fail to receive any of the interventions (IVF or ES) will be excluded from the safety analysis population as advised by the chief investigator. In addition, sensitivity analysis on safety events will be undertaken using treatment assignment as randomised.	Prior to unblinded and blinded review
2.0		Section 11.1 clarifies that women who received any other protocol other than antagonist or long will be excluded in the PP analysis. In addition, we clarified that women who received ES procedure outside the trial will be included in the analysis if they were allocated to the ES procedure	Prior to unblinded and blinded review
2.0		Section 11.6 highlights validity issues of ambiguous pain scores and details how this will be dealt with including additional sensitivity analysis, when appropriate	Pain scores relating to accessibility are recorded in the ES arm only. The change was, therefore, made after unblinded review by the Data Management team and not the blinded Trial Statistician.
2.0		Section 10.14 now includes a clarification that the interest is on unexpected AEs and SAEs that occur after IVF or ES procedure depending on the intervention arm and other circumstances	Prior to unblinded and blinded review.
2.0		Section 11.9.1 clarifies that 95% CI around the difference in proportions will be calculated using Normal approximation to the Binomial distribution.	Prior to unblinded and blinded review.
2.0		Section 11.14.1 now includes a statement to account for the exposure period when modelling the incidences of AEs and SAEs and how the exposure period is calculated.	Prior to unblinded and blinded review.
2.0		Section 11.13 includes an option to use exact methods at the discretion of the Trial Statistician if the observed events are rare.	Prior to unblinded and blinded review.
2.0		Section 11.14.1 reiterates that women who failed to receive any intervention will be excluded in the safety analysis population.	Prior to unblinded and blinded review.
2.0		Section 11.14.2 details relating to the analysis of neonatal SAEs and how centiles of anthropometric measurements will be estimated.	Prior to unblinded and blinded review.

2.1	22/07/2020	Section 11.3 clarifies: 1. the use of FSH (start of stimulation) date when calculating delay in IVF after ES procedure 2. "<" in place of ≤ 37 weeks when classifying a preterm delivery based on gestation age.	During data freeze after unblinding.
2.1		Inserted an appendix in Section 14 with bespoke embryo grading systems developed by an embryologist who was blinded to the trial results. A footnote on Table 7 has been updated to cross reference this appendix. Definitions of acronyms have also been updated to include abbreviations used in this appendix.	During data freeze after unblinding, but the embryologist who developed the grading system was blinded to trial results.

16. References

1 Pye C, Chatters R, Cohen J, *et al.* Induced endometrial trauma (endometrial scratch) in the mid-luteal menstrual cycle phase preceding first cycle IVF/ICSI versus usual IVF/ICSI therapy: study protocol for a randomised controlled trial. *BMJ Open* 2018;**8**:e020755. doi:10.1136/bmjopen-2017-020755

2 ICH. ICH E9: Statistical Principles for Clinical Trials. 1998. http://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Efficacy/E9/Step4/E9_Guideline.pdf

3 Gamble C, Krishan A, Stocken D, *et al.* Guidelines for the Content of Statistical Analysis Plans in Clinical Trials. *JAMA* 2017;**318**:2337. doi:10.1001/jama.2017.18556

4 Sheffield CTRU. ST001: The Statistical Analysis Plan. 2015.

5 Nastri CO, Gibreel A, Raine-Fenning N, *et al.* Endometrial injury in women undergoing assisted reproductive techniques. In: Martins WP, ed. *Cochrane Database of Systematic Reviews*. Chichester, UK: John Wiley & Sons, Ltd 2012. CD009517. doi:10.1002/14651858.CD009517.pub2

6 El-Toukhy T, Sunkara S, Khalaf Y, *et al.* Local endometrial injury and IVF outcome: a systematic review and meta-analysis. *Reprod Biomed Online* 2012;**25**:345–54. doi:10.1016/j.rbmo.2012.06.012

7 Sheffield CTRU. ST007: Randomisation Using the CTRU Randomisation System. 2014.

8 Sheffield CTRU. GV001: Trial Management Group. 2016.

9 Sheffield CTRU. GV002: Trial Steering Committee. 2016.

10 Sheffield CTRU. GV003: Data Monitoring and Ethics Committee. 2016.

11 Sheffield CTRU. DM003: Case Report Form Development. 2016.

12 Sheffield CTRU. DM005: Central Data Validation. 2014.

13 Yelland LN, Sullivan TR, Voysey M, *et al.* Applying the intention-to-treat principle in practice: Guidance on handling randomisation errors. *Clin Trials* 2015;**12**:418–23. doi:10.1177/1740774515588097

14 Moher D, Hopewell S, Schulz KF, *et al.* CONSORT 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials. *BMJ* 2010;**340**:c869. doi:10.1136/bmj.c869

15 Zwarenstein M, Treweek S, Gagnier JJ, *et al.* Improving the reporting of pragmatic trials: an extension of the

CONSORT statement. *BMJ* 2008;**337**:a2390. doi:10.1136/bmj.a2390

16 Senn S. Testing for baseline balance in clinical trials. *Stat Med* 1994;**13**:1715–26.http://www.ncbi.nlm.nih.gov/pubmed/7997705 (accessed 1 Nov 2015).

17 Knol M, Groenwold R, Grobbee D. P-values in baseline tables of randomised controlled trials are inappropriate but still common in high impact journals. *Eur J Prev Cardiol* 2012;**19**:231–2. doi:10.1177/1741826711421688

18 de Boer MR, Waterlander WE, Kuijper LDJ, *et al.* Testing for baseline differences in randomized controlled trials: an unhealthy research behavior that is hard to eradicate. *Int J Behav Nutr Phys Act* 2015;**12**:4. doi:10.1186/s12966-015-0162-z

19 Newcombe RG. Two-sided confidence intervals for the single proportion: Comparison of seven methods. *Stat Med* 1998;**17**:857–72. doi:10.1002/(SICI)1097-0258(19980430)17:8<857::AID-SIM777>3.0.CO;2-E

20 Blizzard L, Hosmer W. Parameter Estimation and Goodness-of-Fit in Log Binomial Regression. *Biometrical J* 2006;**48**:5–22. doi:10.1002/bimj.200410165

21 Kleinman LC, Norton EC. What's the Risk? A Simple Approach for Estimating Adjusted Risk Measures from Nonlinear Models Including Logistic Regression. *Health Serv Res* 2009;**44**:288–302. doi:10.1111/j.1475-6773.2008.00900.x

22 Kahan BC, Morris TP. Improper analysis of trials randomised using stratified blocks or minimisation. *Stat Med* 2012;**31**:328–40. doi:10.1002/sim.4431

23 Cummings P. Methods for estimating adjusted risk ratios. *Stata J* 2009;**9**:175–96.http://ageconsearch.umn.edu/bitstream/127328/2/sjart_st0162.pdf

24 Norton EC, Miller MM, Kleinman LC. Computing adjusted risk ratios and risk differences in Stata. *Stata J* 2013;**13**.https://experts.umich.edu/en/publications/computing-adjusted-risk-ratios-and-risk-differences-in-stata (accessed 29 Jul 2016).

25 Spiegelman D, Hertzmark E. Easy SAS calculations for risk or prevalence ratios and differences. *Am J Epidemiol* 2005;**162**:199–200. doi:10.1093/aje/kwi188

26 Cuzick J. Forest plots and the interpretation of subgroups. *Lancet (London, England)* 2005;**365**:1308. doi:10.1016/S0140-6736(05)61026-4

27 Wang R, Lagakos SW, Ware JH, *et al.* Statistics in medicine--reporting of subgroup analyses in clinical trials. *N Engl J Med* 2007;**357**:2189–94. doi:10.1056/NEJMsr077003

28 Pocock SJ, Assmann SE, Enos LE, *et al.* Subgroup analysis, covariate adjustment and baseline comparisons in clinical trial reporting: current practice and problems. *Stat Med* 2002;**21**:2917–30. doi:10.1002/sim.1296

29 Assmann SF, Pocock SJ, Enos LE, *et al.* Subgroup analysis and other (mis)uses of baseline data in clinical trials. *Lancet (London, England)* 2000;**355**:1064–9. doi:10.1016/S0140-6736(00)02039-0

30 Villar J, Ismail LC, Victora CG, *et al.* International standards for newborn weight, length, and head

circumference by gestational age and sex: The Newborn Cross-Sectional Study of the INTERGROWTH-21st Project. *Lancet* 2014;**384**:857–68. doi:10.1016/s0140-6736(14)60932-6

31 Agresti A. Generalised Linear Models. In: *An Introduction to Categorical Data Analysis*. John Wiley & Sons, Inc 2007. 65–98.