*Early detection of postpartum haemorrhage and treatment using the World Health Organization MOTIVE 'first response' bundle: a cluster randomised trial with health economic analysis and mixed-methods evaluation*

**The E-MOTIVE Trial**

Trial Registration: Clinicaltrials.gov reference number: NCT04341662

# Statistical Analysis Plan

| SAP Version Number | Protocol Version Number |
|---|---|
| *1.0* | *6.0* |

| | | | | | |
|---|---|---|---|---|---|
| Name of Author: | James Martin | Role: | Trial Statistician | Affiliation: | Institute of Applied Health Research<br>University of Birmingham |
| Signature of Author: |  | Date: | 16/03/2023 | | |
| Name of Chief Investigator: | Arri Coomarasamy | Role: | Chief Investigator | Affiliation: | Institute of Metabolism and System Research<br>University of Birmingham |
| Signature of Chief Investigator: |  | Date: | 16/03/2023 | | |
| **This Statistical Analysis Plan has been reviewed and approved by:** | | | | | |

| | | | | | |
|---|---|---|---|---|---|
| Name of Approver: | Lee Middleton | Role: | Senior Statistician | Affiliation: | BCTU<br>University of Birmingham |
| Signature of Approver: |  | Date: | 16/3/2023 | | |
| Name of Approver: | Karla Hemming | Role: | Statistical oversight | | Institute of Applied Health Research<br>University of Birmingham |
| Signature of Approver: |  | Date: | 16/03/2023 | | |

| Abbreviations & Definitions | |
|---|---|
| **Abbreviation / Acronym** | **Meaning** |
| BCTU | Birmingham Clinical Trials Unit |
| CONSORT | Consolidated Standards of Reporting Trials |
| DMC | Data Monitoring Committee |
| ISRCTN | International Standard Randomised Controlled Trial Number |
| ITT | Intention to Treat |
| SAP | Statistical Analysis Plan |
| TSC | Trial Steering Committee |
| TMG | Trial Management Group |
| e-CRF | electronic case report form |
| BERC | Blinded Endpoint Review Committee |
| CRT | Cluster Randomised Trial |
| PPH | Postpartum Haemorrhage |
| | |
| | |
| **Term** | **Definition** |
| International Standard Randomised Controlled Trial Number | A clinical trial registry |
| Protocol | Document that details the rationale, objectives, design, methodology and statistical considerations of the study |
| Randomisation | The process of assigning trial patients or clusters to intervention or control groups using an element of chance to determine the assignments to reduce bias. |
| Statistical Analysis Plan | Pre-specified statistical methods documented for the trial, either in the protocol or in a separate document. |
| | |
| | |
| | |
| | |

# TABLE OF CONTENTS

## 1. Introduction

This document is the Statistical Analysis Plan (SAP) for the E-MOTIVE trial and should be read in conjunction with the current trial protocol. This SAP details the proposed analyses and presentation of the data for the main paper(s) reporting the results for the E-MOTIVE trial.

The results reported in these papers will follow the strategy set out here. Subsequent analyses of a more exploratory nature will not be bound by this strategy, though they are expected to follow the broad principles laid down here. The principles are not intended to curtail exploratory analysis (e.g. to decide cut-points for categorisation of continuous variables), nor to prohibit accepted practices (e.g. transformation of data before analysis), but they are intended to establish rules that will be followed, as closely as possible, when analysing and reporting data.

Any deviations from this SAP will be described and justified in the final report or publication of the trial (using a table as shown in Appendix A). The analyses will be carried out by an appropriately qualified statistician, who should ensure the integrity of the data during the data cleaning processes.

## 2. Background and rationale

The background and rationale for the trial are outlined in detail in the protocol document (current version 6.0). In brief, the study aims to evaluate the implementation of early detection of Postpartum Haemorrhage (PPH) and the use of the WHO MOTIVE 'first response' PPH treatment bundle on clinical, implementation and resource use outcomes. This will be done using a cluster randomised trial design, randomising health facilities to either the E-MOTIVE intervention or to usual care after a baseline phase in which all facilities are subject to usual care.

## 3. Trial objectives

Primary Objective:

Evaluate the implementation of the E-MOTIVE intervention compared with usual care on clinical, implementation and resource use outcomes.

The secondary objectives are as follows:

1. Assess the cost-effectiveness of the E-MOTIVE intervention compared with usual care from a public healthcare system perspective.

2. Develop, optimise and manualise an implementation strategy, with parallel process evaluation alongside the trial, ready for scaling-up of the E-MOTIVE intervention if found to be effective.

The proposed methods for the secondary objectives will be detailed elsewhere.

# 4. Trial methods

## 4.1. Trial design

E-MOTIVE is a prospective, multi-country parallel cluster randomised trial with a baseline control phase (see Appendix B for a schematic of trial design), along with a mixed-methods pilot, process evaluation, and health economic evaluation.

Patients will be included from secondary-level health facilities across the included countries: Kenya, Nigeria, South Africa, and Tanzania.

The unit of analysis, a patient, is defined as a woman who has a verified vaginal birth in a study facility.

A woman with a verified vaginal birth is a woman for whom we have patient and birth information.

In E-MOTIVE, a cluster is defined as a secondary-level health facility. Throughout the rest of this document, the term cluster will be used to represent this.

## 4.2. Trial interventions

**Intervention:** The E-MOTIVE intervention targets healthcare providers and consists of an implementation strategy for early detection of PPH, which allows the initiation of the 'first response' treatment bundle, which is called MOTIVE, consisting of uterine Massage, Oxytocic drugs, Tranexamic acid, IV fluids and Examination & Escalation (See Protocol, Box 2). The timing, dosing and manner with which these interventions will be implemented will be in line with WHO recommendations and local protocols where available. The implementation strategy focuses on (i) calibrated drape with trigger line, (ii) E-MOTIVE emergency trolley and/or carry case, (iii) simulation-based training on-site with peer-assisted learning, (iv) feedback of actionable data to providers, and (v) local E-MOTIVE champions.

**Control:** Usual care with dissemination of the current guidelines. In usual care, 'first response' treatment may include some or all of the components of the MOTIVE treatment

bundle. All components are routinely available in all facilities and recommended for all women with PPH, but currently are used inconsistently by healthcare providers.

## 4.3. Randomisation

After regulatory approvals, 78 clusters from Kenya, Nigeria, South Africa and Tanzania will enter a 7-month baseline period in which they will be following usual care with dissemination of the current guidelines. After this 7-month baseline period, we will allocate 78 clusters (a 1:1 ratio) to either the E-MOTIVE intervention or usual care for 7 months, with an allowance of two months ('transition phase') for full implementation and embedding of the intervention. Clusters not allocated to E-MOTIVE will continue to follow usual care as per the baseline period for the remainder of the intervention phase.

The recruitment of all clusters will be completed within a 3-month period. Following the baseline period, randomisation will be implemented using a minimisation algorithm to ensure a balance of the intervention and control facilities for the following (measured at the cluster-level during the first 5 months of the baseline phase):

1. Number of vaginal births
2. Proportion of births with the composite primary outcome (before randomisation)
3. Oxytocin quality
4. Number of intervention and control clusters in each country

Details of how these are calculated are presented below. The minimisation will be stratified in each country (separate minimisations) to increase the balance in all of the above covariates within a country.

**Number of vaginal births**

After the first five months of the baseline phase has completed, for each cluster, we will gather the total number of births and the number of months of data collection to calculate the cluster's monthly birth rate. We will then use the cluster's monthly birth rate to determine the country's median birth rate. The clusters above and below the country median will form the categories for the minimisation algorithm.

**The proportion of births with the composite primary outcome before randomisation**

After the first five months of the baseline phase has completed, for each cluster, we will gather the total number of births and the total number of events. The ratio of the number of events to the number of births will give us the proportion of births with the composite outcome for that cluster. We will then use the cluster's primary outcome proportion to determine the country's median primary outcome proportion. The clusters above and below the country median will form the categories for the minimisation algorithm.

**Oxytocin Quality**

The quality of oxytocin (measured during the baseline phase) will be dichotomised as ≤90% and >90% to form the categories for the minimisation algorithm.

**Number of intervention and control clusters in each country**

In an attempt to balance the number of clusters allocated to each arm, we will include a covariate that represents the number of clusters assigned to treatment and control, as there is evidence that this improves balance in the number of clusters in each arm [1].

**Minimisation implementation**

The randomisation of health facilities will be performed using a validated minimisation algorithm that follows the Pocock and Simon range method[2] using the "rct_minim" command in Stata 16. The first cluster in each country will be allocated completely at random. All further clusters will be allocated using the minimisation algorithm. To reduce the predictability of allocation and to ensure a balance in covariates across arms, a small random element (10%) will be included in the algorithm. A random element is typically included to better protect the blinding of treatment allocations when recruiting new facilities. A random element of 10% would mean we expect 90% of clusters to be allocated using the minimisation algorithm to the arm that minimises covariate imbalance, and 10% are not.

The randomisation of health facilities will be performed using a validated minimisation algorithm and implemented by an independent statistician from the University of Birmingham Clinical Trials Unit (BCTU). Full details of the randomisation specification will be stored in a confidential document at the University of Birmingham Clinical Trials Unit (BCTU).

## 4.4. Timing of outcome assessments

All outcomes are measured between birth and discharge from the cluster or higher-level facility (if transferred).

## 4.5. Primary outcome measure

The primary outcome is a composite of the following three clinical outcomes:

- Primary severe PPH (defined as blood loss ≥1000 ml) following vaginal birth in the cluster measured up to 2 hours postpartum.
- Postpartum laparotomy for bleeding until discharge from the cluster.
- Postpartum maternal death from bleeding until discharge from the cluster.

If any of the components occur, this will be deemed as positive for the primary outcome.

The components of the composite primary outcome will be treated as secondary outcomes and examined individually as is recommended [3].

**Measurement of blood loss**

Blood loss will be collected using blood collection drapes for all women after vaginal birth. All facilities will use non-calibrated drapes during the baseline phase. During the intervention phase, clusters allocated to the control arm will continue to use non-calibrated drapes and clusters allocated to the intervention arm will use calibrated drapes containing measurement and trigger lines. The calibrated drapes are identical to the non-calibrated drapes with the only difference of having measurement and trigger lines to aid healthcare providers detect and treat PPH early. Blood loss will be measured for all women by weighing the blood collection drapes at the first hour postpartum, or if the bleeding continues, the blood collection will be extended to the second hour postpartum. The weight of the blood collection drape containing the blood will be recorded using digital weighing scales. If there is excessive, additional blood loss beyond the second hour postpartum up to 24 hours postpartum a visual estimation of blood loss will also be provided, and this will be included in the overall blood loss measurement as a secondary clinical outcome.

Only blood loss data that has been source verified will be included in the primary analysis – see below for details. Blood loss data that has not been source verified is included in a sensitivity analysis.

**Source-verified data**

Analyses of outcomes with blood loss data will only include blood loss that has been source-verified. Blood loss measured using the blood collection drape, which is source-verified will have a corresponding photograph, which is uploaded to a secure portal, of the drape on the E-MOTIVE digital weighing scale displaying the drape weight. The photograph should also display the drape weight documented on the E-MOTIVE stickers before being placed in the patient notes. The drape weight value entered on to the E-MOTIVE REDCap database is verified against the drape weight in the photograph and the photograph is linked to the patient record on the database using the E-MOTIVE patient I.D number.

**Blinded Endpoint Review Committee**

A Blinded Endpoint Review Committee (BERC) will assess incoming data relevant to the primary outcome to confirm if any postpartum laparotomy was performed for bleeding and if any maternal death was due to bleeding. Initially, two central independent reviewers (identified by the TMG) will review de-identified data via the online E-MOTIVE database where there is a case of postpartum laparotomy and/or maternal death. The reviewers will not be able to view the facility's responses as to whether the postpartum laparotomy was for bleeding and/or the maternal death was due to bleeding. Each reviewer will be prompted to log in to the E-MOTIVE database and record their assessment on an electronic case report form (e-CRF) following receipt of an automated notification. If there is discordance between the assessments of the two reviewers or if their assessment conflicts with the assessment provided by the facility then a further assessment will be required by a central independent arbitration committee of five members convened by the TMG. The committee will meet as frequently as required and a decision will be made by a majority vote. The arbitration

committee will have access to the facility assessment (on the Discharge Outcome form) and the two independent reviewer assessments in order to make their decision. The Programme Manager (or delegate) will minute committee discussions and enter a summary of the discussions along with their decision on to the e-CRF. The review by the BERC members will be undertaken blinded to the randomised allocation of the health facility. Members of the BERC are required to formally register their assent to join the committee by signing a separate charter. Further details of the BERC can be found in their charter.

## 4.6. Secondary clinical and implementation outcome measures

**Key Implementation Outcomes**

The key implementation outcomes are:

1) Postpartum haemorrhage detection
2) Compliance with the MOTIVE bundle

These are defined below.

**Postpartum haemorrhage detection**

Postpartum haemorrhage (PPH) detection has the following numerator and denominator: women who objectively had PPH (source verified blood loss ≥ 500 mL after weighing the drape) and were diagnosed with PPH by the birth attendants divided by the total number of women who objectively had PPH (source verified blood loss ≥ 500 mL after weighing the drape).

**Compliance with the MOTIVE bundle**

Compliance with the MOTIVE bundle has the following numerator and denominator: women who objectively had PPH and were treated with the PPH bundle following a diagnosis of PPH by the birth attendants divided by the total number of women who objectively had PPH (blood loss ≥ 500 mL after weighing of the drape). Compliance with the MOTIVE bundle is defined as adherence to the three core elements of the bundle: administration of oxytocic drugs, TXA and IV fluids. If all three core elements are administered when a PPH is diagnosed, this will be deemed positive for bundle compliance. If any of the three core elements are not administered when a PPH is diagnosed, then this will be deemed negative for bundle compliance.

**Secondary Clinical Outcomes**

The secondary clinical outcomes are based on the Core Outcome Set for PPH treatment, and are the following:

1. Laparotomy postpartum until discharge from the cluster*;
2. Laparotomy with compression sutures postpartum until discharge from the cluster*;
3. Laparotomy with arterial ligation postpartum until discharge from the cluster*;
4. Hysterectomy postpartum until discharge from the cluster*;
5. Hysterectomy postpartum for bleeding until discharge from the cluster*;
6. All-cause maternal mortality postpartum until discharge from the cluster*;
7. Blood loss (reported in ml) up to 24 hours postpartum†;

8. Primary PPH defined as blood loss ≥500 ml up to 24 hours postpartum†;
9. Duration of hospitalisation postpartum;
10. Duration of ICU hospitalisation postpartum until discharge from the cluster†;
11. Transfers to higher-level facility postpartum until discharge from the cluster;
12. All-cause neonatal mortality postpartum until discharge from the cluster*;
13. Use of Non-pneumatic anti-shock garment (NASG) postpartum†;
14. Use of uterine balloon tamponade postpartum until discharge from the cluster†;
15. Blood transfusion postpartum until discharge from the cluster†*;
16. Blood transfusion for postpartum haemorrhage until discharge from the cluster†*;
17. Intensive Care Unit (ICU) admissions postpartum until discharge from the cluster†*;
18. Primary severe PPH (defined as blood loss ≥1000 ml) following vaginal birth in the cluster measured up to 2 hours postpartum†;
19. Postpartum laparotomy for bleeding until discharge from the cluster*;
20. Postpartum maternal death from bleeding until discharge from the cluster*;

†Combined clinical and quality of care exploratory outcomes where we may observe an increase or a reduction if E-MOTIVE is effective.
*In cases where a woman is transferred to another facility postpartum, discharge from the cluster relates to the facility the woman is transferred to and the relevant outcomes will be obtained from the facility to which the woman was referred.

**Secondary Implementation Outcomes**
The secondary implementation outcomes are:
1. PPH treatment (with the following numerator and denominator: women diagnosed with PPH by the birth attendants divided by the total of women having a vaginal birth in the health facility);
2. Bundle usage (with the following numerator and denominator: women treated with the PPH bundle following a diagnosis of PPH by the birth attendants divided by the total of women having a vaginal birth in the health facility);
3. Bundle usage for PPH (with the following numerator and denominator: women treated with the PPH bundle following a diagnosis of PPH by the birth attendant divided by the total of women diagnosed with PPH by the birth attendants);
4. Uterine massage;
5. Oxytocin use;
6. Misoprostol use;
7. TXA use;
8. Intravenous fluids use;
9. Examination of the genital tract;
10. Number of women receiving any treatment uterotonic;
11. Number of women requiring additional treatment interventions (not responding to the MOTIVE bundle).

All secondary implementation outcomes will be reported with three denominators:

- The total study population.
- Women diagnosed with PPH by the birth attendant.
- Women who objectively had PPH (blood loss ≥ 500 mL after weighing the drape).

## 4.7. Sample size

In October 2021, a revised sample size calculation was performed, based on the study shortening from 22 months to 14 months (plus a 2-month transition period). As part of this, using data from the accrued baseline period (collected from October 2020 to September 2021), we estimated some of the parameters assumed in the earlier sample size and power calculations. The revised sample size calculation was made based on empirical baseline data from clusters being included in the trial.

Below, we describe the methods used to calculate the sample size.

The original estimates of the parameters required for the sample size calculation are described, alongside updated estimates based on the data from the baseline period. Full details of the original sample size calculation can be found in the protocol.

Assuming there are 80 health facilities in the trial, evenly split across the intervention and control groups, with an average number of 192 births per cluster per month, the anticipated total sample size for the study (running for 14 months) would be 215,040 (=80*192*14). The number of health facilities (80) has been inflated by 10% to allow for dropout from the number of health facilities required (72).

In the control condition, we initially expected the prevalence of the primary composite outcome to be around 2%. This prevalence is informed by the prevalence of a similar outcome based on the analysis of the CHAMPION trial, a large multi-country study of PPH prevention[4]. We considered sensitivity to this prevalence between 1.5% and 4%. Analysis of the baseline data from E-MOTIVE clusters, however, found a prevalence of the primary outcome of 3.6% (95% CI: 3.4% to 3.8%).

A relative risk reduction of 25% was considered to be a clinically meaningful difference to detect, though smaller reductions could also be worthwhile. Under some scenarios (such as a high value of the ICC) this target effect size might not be detectable at 90% power. We have thus considered sensitivity to the power for a relative reduction of 20% and 30%.
Sample size calculations have allowed for the clustered nature of the design through the intra-cluster correlation (ICC). As recommended, values for the ICC have been informed through a combination of the literature and analysis of available data on a similar set of outcomes. Analysis of the CHAMPION trial found an ICC of 0.03 (95% CI: 0.02 to 0.05) for a similar composite outcome of PPH ≥1000 ml, laparotomy and maternal death. This ICC was estimated using a linear mixed regression model and so is on the proportions scale as is appropriate for a sample size calculation. As the literature suggests, rare clinical outcomes tend to have small ICCs, for this reason, we have considered an extended range for the ICC,

from 0.001 to 0.05, with an expected ICC of 0.02. Analysis of the baseline data from the E-MOTIVE clusters found an ICC of 0.018 (95% CI: 0.01 to 0.034).

To allow for variations in clustering over time, we have allowed for a cluster-by-period random effect. This has been incorporated in the sample size calculations using the cluster autocorrelation (CAC). There is limited information from the literature on the likely values of the CAC. Using data from the CHAMPION trial, we created two 7-month periods, which match the planned study design. From this, we estimated a CAC of 0.97. No current methodology exists to calculate a confidence interval for this value. We have chosen to consider sensitivity to the CAC between the values of 0.95 and 1.0. No estimate of the CAC could be made using the accrued baseline data.
Appendix Table E7 provides information on the expected levels of power for a 14-month study (two 7-month periods) across the range of parameters described above. These suggest that the study has at least 90% power at 5% significance (two-sided) to detect a 30% RRR for most scenarios after allowing for clustering and varying cluster sizes. The study will have over 90% power to detect smaller RRR if the ICC is close to the lower bound (0.001), the CAC is at the upper bound (1.0), or the prevalence of the study is relatively large (4.0%).
With large prevalence (4.0%), the study would have over 90% power to detect a 20% RRR for most likely scenarios (Appendix E7). A re-estimation of the CAC using two 7-month periods in the CHAMPION trial data found a CAC of 0.99. With a CAC of 1.00, the study would have over 90% power to detect a 20% RRR for most likely scenarios.

## 4.8. Interim analyses and stopping guidance

Arrangements for Data Monitoring Committee (DMC) oversight has been agreed, and details can be found in the DMC Charter.

## 4.9. Internal Pilot Progression Rules

Not Applicable

## 4.10. Timing of final analysis

The final analysis for the trial will occur after all clusters have been randomised, ran for a fixed period of time post-randomisation (7-months, and the corresponding outcome data have been entered onto the trial database and validated as being ready for analysis. This is provided that the trial has not been stopped early for any reason (e.g., DMC advice to TSC or funding body request).

## 4.11. Timing of other analyses

Not applicable

## 4.12. Trial comparisons

All references in this document to 'group' refer to E-MOTIVE or usual care.

# 5. Statistical Principles

## 5.1. Confidence intervals and p-values

All estimates of differences between groups will be presented with two-sided confidence intervals of appropriate size.

As described earlier, the outcomes can be grouped in to four categories: the primary outcome; the key implementation outcomes; secondary clinical outcomes; and secondary implementation outcomes. See section 4.5 and 4.6 for details of the outcomes. P-values will be reported for the primary outcome only. For all other outcomes (key implementation, secondary clinical and secondary implementation), we will not present p-value unless requested to by a peer reviewer. To this end, summary measures with 95% confidence intervals will be presented as supporting evidence only, and should be cautiously interpreted due to the possibility of multiplicity. For clarity, the information presented for each outcome group is highlighted below.

**Primary Outcome**

We will report the number of patients in each group (E-MOTIVE or usual care) stratified by time-period, the summary measure, and a two-sided 95 % confidence interval. Hypothesis testing will be carried out at the (two-sided) 5% level of significance. A p-values will be reported for the primary outcome.

**Key implementation outcomes**

We will report the number of patients in each group (E-MOTIVE or usual care) stratified by time-period, the summary measure, and a two-sided sided 95 % confidence interval. We will not report p-values for secondary outcomes, unless requested by external statistical reviewer(s).

**Secondary clinical outcmes**

We will report the number of patients in each group (E-MOTIVE or usual care) stratified by time-period, the summary measure, and a two-sided sided 95 % confidence interval. We will not report p-values for secondary outcomes, unless requested by external statistical reviewer(s).

**Secondary implementation outcomes**

We will report the number of patients in each group (E-MOTIVE or usual care) stratified by time-period, the summary measure, and a two-sided sided 95 % confidence interval. We will

not report p-values for secondary outcomes, unless requested by external statistical reviewer(s).

### 5.2. Adjustments for multiplicity

No correction for multiple testing will be made.

For all secondary clincal and implementation outcomes (see section 4.6), results will be interpreted with caution due to the potential for multiplicity issues.

### 5.3. Analysis populations

All primary analyses (primary and secondary outcomes including safety outcomes) will be by intention-to-treat (ITT) with appropriate adjustment for clustering. Patient and clusters will be analysed in the group to which they were randomised, and all patients (and clusters) shall be included whether or not they received the allocated intervention.

## 6. Trial population

### 6.1. Cluster Inclusion Criteria

Health facilities (clusters) are eligible for inclusion if they have 1000 to 5000 births a year and provide comprehensive obstetric care with ability to perform surgery for PPH. Clusters are selected based on being administratively and geographically distinct from each other. Pre-existing implementation of early detection or bundled approach for PPH management are exclusion criteria.

The cluster start data for inclusion in the study, is the date in which ≥80% of their patients have source data verification.

### 6.2. Research Participant Inclusion Criteria

All healthcare providers attending vaginal births at the study facilities.

### 6.3. Patient Inclusion Criteria

All women who have a verified vaginal birth in the included health facilities. A woman with a verified vaginal birth is a woman for whom we have patient and birth information.

## 6.4. Cluster and participant flow

A flow diagram (as recommended by CONSORT[5]) will be produced to describe the participant and cluster flow through each stage of the trial. An example of how this may look is presented in Appendix E1. In this trial, clusters are recruited, but there is no participant recruitment (all eligible patients are included in the study).

**Cluster flow**

We will present the number of clusters recruited, the number of clusters that were randomly assigned, the number (with reasons) of any clusters lost to follow-up (drop-outs and withdrawals), and the number that were analysed for the primary outcome and key implementation outcomes.

**Participant flow**

We will present the number of patients meeting the inclusion criteria in the baseline period and the post-randomisation period, as well as the number of patients that were analysed for the primary outcome and key implementation outcomes.

CRTs are prone to a variety of biases, depending on recruitment processes, timing of recruitment, and whether blinding occurs. To improve clarity of the timing of the trial processes and blinding, which are not covered by reporting guidelines, we will use the timeline cluster graphical tool to clarify the processes used here and to identify potential bias in the study[6]. An example is presented in Appendix E2.

## 6.5. Baseline characteristics

Characteristics of the health facilities and the individual patients will be summarised as per Appendix E3.

We will tabulate the characteristics of the health facilities, stratified by allocation to E-MOTIVE or control, and stratified by time-period (baseline or post-randomisation), and present the number of facilities in each group, with percentages.

We will tabulate the characteristics of included patients, stratified by allocation to E-MOTIVE or control, and stratified by time-period (baseline or post-randomisation). Categorical data will be summarised by number of patients, counts and percentages. Continuous data will be summarised by the number of patients, mean and standard deviation if deemed to be normally distributed, or number of patients, median and interquartile range if data are skewed, and ranges if appropriate. For the main trial report, tests of statistical significance will not be undertaken, nor confidence intervals presented[7], unless requested by external statistical reviewer(s).

# 7. Analysis methods

The study's primary aim is to evaluate whether there is a difference in the primary composite outcome in the intervention group (E-MOTIVE) compared to the control group.

Below, we present the model choice and analysis methods for the primary outcome, the secondary clinical outcomes and the secondary implementation outcomes. We present methods for accounting for clustering and methods for dealing with model non-convergence. We present the plan for adjusting for covariates and describe the sub-group and sensitivity analyses.

In summary, we will be comparing E-MOTIVE and the control using a constrained baseline analysis, as recommended by Hooper et al[8]. Mixed-effect regression models are fitted to the data, with a random effect to allow for clustering, and random cluster-period effect to allow the correlation between observations within a cluster to vary if the observations are made at different time periods. Fixed effects are included for time-period, allocated exposure to E-MOTIVE, and covariates used in the randomisation. The covariate representing allocated exposure to E-MOTIVE will indicate the impact of E-MOTIVE.

The analysis of the primary outcome is described in sections 7.4. The analysis of secondary clinical and implementation outcomes is described in sections 7.5.

See Appendix D for information on how variables will be derived for the analysis. A template for reporting the primary outcome is given in Appendix E4, and the secondary clinical and implementation outcomes in E5.

## 7.1. Estimands

Since E-MOTIVE is a cluster trial, the summary measure of the effect of E-MOTIVE can be measured in different ways. That is to say, the unit of inference of whether the treatment is beneficial can be made at the unit of randomisation (cluster) or at the unit of data collection (participant). Please see Appendix E8 for further information.

In E-MOTIVE, for all outcomes (both clinical and implementation), we will be estimating the participant-average summary measure. This will indicate the average summary measure across patients, that is, how effective E-MOTIVE is for the average participant. See sections 7.4 for details on the analysis methods.

The list of clinical outcomes and implementation outcomes is given in section 4.6.

## 7.2. Informative cluster sizes

In cluster trials, the choice of analysis method is influenced by whether there are informative cluster sizes – which is when the size of a cluster is associated with the outcome in that cluster[9]. This could be because the outcome rate differs between small and large clusters, or because the summary measure differs between small and large clusters. In E-MOTIVE, it is plausible that there may be informative cluster sizes, and the analysis methods proposed below contain an adjustment for cluster size, in an aim to account for this. We will also fit generalised estimating equations with independent working correlation matrix as a sensitivity analysis (see section 7.11)[9].

## 7.3. Summary Measures

For all binary outcomes – this includes the primary composite outcome, the key secondary implementation outcomes, and other secondary clinical and implementation outcomes, we will summarise the number of patients who experience the outcome, and the percentage of the group this corresponds to. Summary statistics will be presented stratified by trial arm (E-MOTIVE or control) and time-period (baseline phase or post-randomisation phase). To quantify the effect of E-MOTIVE, the intervention effect will be presented using relative risk and risk differences – as per CONSORT recommendations[3].

For secondary outcomes that are continuous, we will summarise using means and standard deviations or medians and inter-quartile ranges as appropriate. Histograms will be produced to determine which summary statistic is most appropriate. We will present summary statistics stratified by trial arm (E-MOTIVE or control) and time-period (baseline phase or post-randomisation phase). To quantify the effect of E-MOTIVE, the intervention effect will be presented using the mean difference (difference in means between E-MOTIVE and usual care).

## 7.4. Primary outcome analysis

The model being fitted to the binary primary composite outcome will be a mixed-effect logistic regression model with robust standard errors.

To account for clustering, random effects (for cluster and cluster-period) are included (see section 7.4.1). We will account for the time-period using fixed-effects (see section 7.4.2), and will adjust for covariates used in the randomisation in the first instance (see section 7.4.3). If there are issues with model-convergence, then a plan is in place (see section 7.4.4).

Risk ratios and risk differences will then be derived using marginal standardisation [10], as in for example a CRT by Kirkwood et al [11]. Under this approach, the mean risk under the control condition and intervention condition is obtained. The risk ratio is then derived using the mean risk in intervention condition and by the mean risk in the control condition, with the calculation

being performed on the log scale. The risk difference is derived as the mean risk in intervention condition minus the mean risk in the control condition. We will use the unconditional standard errors to obtain the confidence interval, as this allows for correlation amongst the observations [12].

A small sample correction will not be included in the main analyses as the number of clusters is expected to be sufficiently large to maintain the type I error rate [13].

The integration method used will be mean-variance adaptive Gauss-Hermite quadrature, which is the default in Stata.

### 7.4.1. Accounting for clustering of observations

To allow for correlation between observations within a cluster, we will fit mixed-effect logistic regression models with robust standard errors. Clustering will be allowed for in all analyses through the use of random effects. Recent developments in the methodological literature for the design and analysis of multiple-period cluster randomised trials suggest that a simple exchangeable correlation structure is not sufficient to depict the correlation structure in multiple-period cluster randomised trials [8]. We will include a random cluster by period interaction in addition to a random cluster effect to allow for the decay in correlation between observations made in the same cluster, but at different time periods.

### 7.4.2. Primary outcome - accounting for baseline period

The analysis will use a constrained baseline analysis approach, as recommended by Hooper et al[8]. Here, outcomes collected during the baseline period and the post-randomisation period are treated as longitudinal and a repeated measures analysis is used to estimate the summary measure.

Under this approach, there are two binary covariates included as fixed effects in the model: (i) period and (ii) exposure to E-MOTIVE. These covariates are described below.

**Covariate for period**
A binary variable will indicate whether observations are taken from the baseline period (0) or the post-randomisation period (1) of the study.

**Covariate for exposure to E-MOTIVE**
A binary `treatment' variable will indicate whether observations are exposed to the E-MOTIVE bundle. That is, observations taken in the post-randomisation period in clusters allocated to E-

MOTIVE will have a value of 1, and all other observations (control clusters and E-MOTIVE clusters during baseline phase) will have a value of 0. This binary treatment variable will be used to determine the impact of the intervention. Under this approach, it is assumed that there is no systematic differences between the groups at baseline.

### 7.4.3. Primary outcome – covariate adjustment

In the first instance, the analysis will contain an adjustment for the minimisation parameters listed in section 4.3. Randomisation was performed stratified by country, and so country will also be included as a covariate. However, whilst the number of intervention and control clusters was included in the minimisation, we will not adjust for the number of clusters allocated to E-MOTIVE or usual care.

The number of vaginal births will be modelled as binary (above or below median). Oxytocin quality will be treated as binary (high quality or low quality). Country will be treated as categorical (Kenya, Nigeria, South Africa, or Tanzania).

For the analysis of the primary outcome, we will not include a term for cluster-level primary outcome rate at baseline, as although this was included as a randomisation covariate, the constrained baseline analysis method includes individual-level primary outcome values from the baseline period.

### 7.4.4. Primary outcome - model non-convergence

If the mixed-effect logistic regression model with random cluster and cluster-period effects does not converge, we will remove the cluster-period random effects and fit a mixed-effects model with a random effect for cluster.

If this model does not converge, then we will perform an cluster-level analysis weighted by cluster size to ensure we are targeting the participant-level effect [9]. To perform a cluster-level analysis, for each outcome and for each cluster we will calculate a cluster-level summary statistic for the baseline period and for the post-randomisation period separately. The cluster-level summary statistic will be the proportion of the cluster who experience the outcome.

The data will then be transformed from long format to wide format. That is, instead of having two observations per cluster (one observation that describes the summary statistic during the baseline period and one observation that describes the summary statistics during the post-randomisation phase), each cluster will have one row of data.

**Relative risk**

To obtain a relative risk, we will fit a linear regression model to a transformed cluster-level summary statistic (log transformation) with a fixed effect for allocated exposure to E-MOTIVE and a fixed-effect for baseline summary measure. We will take the exponential of the covariate related to allocated exposure to E-MOTIVE to obtain the relative risk and 95% confidence intervals. We will use the unconditional standard errors to obtain the confidence interval, as this allows for correlation amongst the observations [12].

**Risk difference**

To obtain the risk difference, a linear regression model will be fitted to the cluster-level proportions with a fixed-effect for allocated exposure to E-MOTIVE and a fixed-effect for baseline summary measure. We will use the unconditional standard errors to obtain the confidence interval, as this allows for correlation amongst the observations [12].

## 7.5. Analysis methods – secondary outcomes

See Appendix D for information on how variables will be derived for the analysis. A template for reporting the secondary outcomes is given in Appendix E5.

**Binary outcomes**

All secondary clinical outcomes that are binary will be analysed in the same way as the primary outcome (see section 7.4). In summary, they will be analysed using a mixed-effect logistic regression model, with a random effect for cluster and a random effect for cluster-by-period. The analysis will use a constrained baseline analysis. We will adjust for covariates used in the randomisation process: country (Kenya, Nigeria, South Africa, or Tanzania); Oxytocin quality (high quality or low quality); number of vaginal births (above median or below median); and primary outcome prevalence (above median or below median). Risk ratio and risk differences will be obtained from the logistic model using marginal standardisation. If these models fail to converge, we will remove the cluster-period random effects and fit a mixed-effects model with a random effect for cluster. If this model also fails to converge, an unweighted cluster-level analysis will be performed in the manner described in section 7.4.4.

**Continuous outcomes**

All secondary clinical outcomes that are continuous will be analysed using a linear mixed-effect model. The analysis will use a constrained baseline analysis, in the same format as the primary outcome. A binary variable will indicate whether observations are taken from the baseline period (0) or the post-randomisation period (1) of the study. A binary `treatment' variable will indicate whether observations were exposed to the E-MOTIVE intervention. That is, observations taken in the post-randomisation period in clusters allocated to E-MOTIVE will have a value of 1, and all other observations (control clusters and E-MOTIVE clusters during baseline phase) will have a value of 0. This binary treatment variable will be used to determine the impact of the intervention. To allow for correlation of observations within a cluster, a random effect for cluster

will be included in the model. An adjustment will be made for the period effect to vary across clusters (cluster by period random effect). We will adjust for covariates used in the randomisation process: country (Kenya, Nigeria, South Africa, or Tanzania); Oxytocin quality (high quality or low quality); number of vaginal births (above median or below median); and primary outcome prevalence (above median or below median).

If the linear model with random cluster and random cluster by period effects fails to converge, we will remove the cluster-period random effects and fit a mixed-effects model with a random effect for cluster.

If this model also fails to converge, we will fit an unweighted cluster-level analysis. Under this approach, for each cluster we will calculate a cluster-level summary statistic for the baseline period and for the post-randomisation period separately. This will be the mean value of the outcome in that cluster in that period.

The data will then be transformed from long format to wide format. That is, instead of having two observations per cluster (one observation that describes the summary statistic during the baseline period and one observation that describes the summary statistics during the post-randomisation phase), each cluster will have one row of data.

To obtain the mean difference, a linear regression model will be fitted to the cluster-level mean with a fixed effect for allocated exposure to E-MOTIVE and a fixed-effect for cluster-level mean at baseline. Regression models will be fitted unweighted, as models with size-weighting and variance-weighted have been shown to have lower type 1 error rates [13, 14].

## 7.6. Distributional assumptions and outlying responses

Distributional assumptions (e.g., normality of regression residuals for continuous outcomes) will be assessed visually prior to analysis. In the first instance the proposed primary method of estimation in this analysis plan will be followed. If responses are particularly skewed and/or distributional assumptions violated, the impact of this will be examined through sensitivity analysis; this will consist of transformation of responses prior to analysis (e.g., log transformation) in the first instance.

If extreme values are apparent and considered to be affecting the integrity of the analysis, a sensitivity analysis consisting of removing the outlying response(s) and repeating the analysis will be performed.

Output from these analyses, if performed, will be described, and presented alongside the original analysis (or included, e.g., in appendices) with the excluded values clearly labelled. See section 7.11 for further details regarding sensitivity analyses.

## 7.7. Handling missing data

In the first instance, analysis will be completed on received data only with every effort made to follow-up patients to minimise any potential for bias. In E-MOTIVE, there is the potential for missing outcome data and missing covariate data, which is described below. A planned sensitivity analysis using multiple imputation to allow for missing data is described in section 7.11.

**Missing outcome data**

**Primary outcome**

Only blood loss data that is source verified is included in the primary analysis of the primary outcome. There are two reasons that a woman could having missing outcome data, (i.e., missing source verified blood loss data): 1) woman has blood loss data recorded but it is not source verified; and 2) woman has no blood loss data recorded on the case report form.

**Secondary clinical outcomes**

Secondary clinical outcomes may contain some missing data. Blood loss outcomes (blood loss 2 hours postpartum and 24 hours postpartum) and primary PPH may contain missing data for the same reason as the primary outcome. Other outcomes may contain missing information if the case report form containing outcome data has not been completed fully.

**Secondary Implementation outcomes**

For all implementation outcomes, the case report form denotes whether a woman received the outcome in question. It is assumed if it was not denoted as a "yes" on the caser report form that the woman did not receive the outcome. As such, there is no missing data for the implementation outcomes.

**Missing covariate data**

All cluster-level data included in (unadjusted) primary analysis models (country, oxytocin quality, cluster size) will be complete and not contain any missing information. Country is known for each participant based on their cluster. The quality of Oxytocin has been measured and evaluated for all clusters. The cluster size will be based on the number of patients in the cluster during the baseline period.

Individual data included in the (adjusted) sensitivity analysis models (age; parity; multiple pregnancy; mode of birth; and birth weight) may contain missing data.

## 7.8. Analysis methods – exploratory outcomes and analyses

Not applicable

## 7.9. Safety data

Analysis of all outcomes including safety outcomes are covered in section 7.10.

## 7.10. Planned subgroup analyses

We will estimate and report subgroup effects irrespective of their statistical significance. The results of subgroup analyses will be treated with caution. Analysis will be limited to the primary composite outcome and key implementation outcomes only. The planned subgroup analyses are:

- Baseline period PPH rate (above median vs below median) as objectively diagnosed by blood collection drape using data from the baseline period
- Number of vaginal births (using baseline period size) (above median size vs below median size)
- Age (three levels categorised by tertiles of the distribution)
- Parity (none vs any)
- Type of pregnancy (singleton vs multiple),
- Gestational age (term vs preterm),
- Mode of birth (spontaneous vs assisted),
- Presence of any risk factors for PPH vs no risk factors for PPH (Risk factors for PPH are: diagnosis of preeclampsia, induction or augmentation of labour, previous caesarean, previous post-partum haemorrhage, antepartum haemorrhage, episiotomy, vaginal tears, retained placenta)
- Country (Kenya vs Nigeria vs South Africa vs Tanzania).
- Compliance with the bundle (facilities split into above or on median vs below median). Compliance with the MOTIVE bundle is defined as adherence with three core elements of the bundle: administration of oxytocic drugs, TXA and IV fluids. If all three core elements are administered, this will be deemed positive for bundle compliance.
- Time since intervention exposure (exposure (defined as number of months since cluster began post-randomisation phase) (described below)

**Time by treatment interaction**

To evaluate whether the intervention effect is consistent over time, we will estimate the time by treatment interactions. To do this, we will categorise observations made in the post-randomisation phase to: month 1; month 2; month 3; month 4; month 5; month 6; and month 7.

The analysis of the subgroups will be performed by including a treatment by subgroup interaction parameter in the regression model. Subgroup analyses that are at the cluster level (for example, country) will contain a between-within small sample correction – which is recommended when using mixed models for binary outcomes[15]. For this, we will use the t-

distribution with the appropriate degrees of freedom (number of clusters minus number of parameters estimated in the model).

For each subgroup analysis listed above, we will report the summary measure with 95% confidence interval in each sub-group and the p-value for interaction parameter. We also plan to present the relative measure of differences (ratio of ratios) between subgroup-specific summary measures, along with measures of uncertainty. The "ratio of ratios" will be obtained using post-estimation commands.

All sub-groups are for checking for consistency across various levels of the subgroup.

A template for reporting the subgroup analyses for the primary outcome is given in Appendix E6. These results may be presented as a forest plot to aid interpretability of the results.

## 7.11.Supportive and sensitivity analyses

Sensitivity analyses will be limited to the primary composite outcome and key implementation outcomes and will consist of:

- Fully adjusted analysis
- Accounting for Missing Data (Missing at Random)
- Accounting for Missing Data (Missing Not at Random)
- Independent estimating equations (Informative clustering)

**Fully adjusted Analysis**

For the adjusted analysis, we will use a propensity score (PS) based method – the direct PS adjustment as recommended by Leyrat et al[16]. Under this method, the PS is obtained by fitting a multivariable logistic model with potential confounders (age; parity; multiple pregnancy; mode of birth; and birth weight) in the model, and treatment arm as the outcome. The PS is then the probability of being in the E-MOTIVE arm given these observed confounders. Clustering is not accounted for in the estimation of the PS as treatment arm is nested within cluster. Age and birthweight will be included as continuous covariates. For continuous covariates, we will model the relationship using fractional polynomials if they better depict the relationship than a linear trend, with a $p<0.05$ threshold. Including continuous covariates with fractional polynomials is recommended when the relationship between covariate and outcome is not known [17]. Parity (none vs any), multiple pregnancy (singleton vs multiple), and mode of birth (spontaneous vs assisted) will be treated as binary. We will constrain the propensity

score for baseline observations to be zero - since no patient can receive intervention in that time period.

In the direct adjustment approach, the PS is then included in the analysis model as a covariate in its continuous form. We will include the PS with fractional polynomials if this better predicts the relationship with the outcome than including the PS in a linear form. We will use a $p<0.001$ threshold for including fractional polynomial form.

The model fitted to the data for the sensitivity analysis will therefore include the minimisation parameters (as discussed in *Primary Analysis*, above), and a covariate for the propensity score.

As agreed in the DMC charter, at the 3rd DMC meeting (after approximately 75% of births), the DMC will be asked to consider whether there is any clear evidence of substantial selection bias. In the unlikely event that substantial selection bias is confidently identified, then the TSC and TMG will consider whether the sensitivity analysis (with PS) will become the primary analysis for all outcomes. Any decision will be documented in the meeting minutes.

**Accounting for missing data**

The potential for missing outcome data and missing covariate data is discussed in section 7.7. Below, we present multiple imputation methods to deal with missing outcome and individual-level covariate data. We consider this under two assumptions:

1) Data missing is missing at random (MAR)
2) Data missing is missing not at random (MNAR).

To evaluate the impact of missing data under MAR and MNAR mechanisms, multiple imputation will be used. We will impute the primary outcome (for all women who do not have source-verified primary outcome data) and individual level covariates (for any covariate used in the adjusted analysis).

**Missing at Random**

We firstly perform multiple imputation using chained equations [18]. The primary outcome and all covariates required for the adjusted analysis with missing data will be included in the imputation process. The following auxiliary variables will be included to aid the imputation process: country; cluster size (above or below median); Oxytocin quality (high or low); time period (baseline or post-randomisation); exposure to E-MOTIVE; previous C-section; gestational age; antepartum haemorrhage; retained placenta; and non-source verified blood loss.

The number of imputations, *n*, will be equal to the proportion of patients with any missing primary outcome data. To ensure inferences are valid, clustering will be allowed for in the imputation process. This will be done using BLIMP 3.0 which allow a random effect for cluster to be included [19].

In each imputed dataset, a propensity score is obtained by fitting a multivariable logistic model with potential confounders (age; parity; multiple pregnancy; mode of birth; and birth weight) in the model, and treatment arm as the outcome. In each imputed dataset, the analysis model with PS included in the model as a covariate is fitted and an estimate of the summary measure is obtained. The RR and RD will be obtained in the same manner as for the primary analysis – using marginal standardisation. The summary measure from each of the *n* imputed datasets is then pooled using Rubin's rules.

**Missing Not at Random**

An analysis to assess the effect of missing responses for the primary outcome only. This analysis will explore the possibility that missing responses are 'missing not at random' (MNAR) using a tipping point approach.

We firstly perform multiple imputation using chained equations [18]. The primary outcome and all covariates required for the adjusted analysis with missing data will be included in the imputation process. The following auxiliary variables will be included to aid the imputation process: country; cluster size (above or below median); Oxytocin quality (high or low); time period (baseline or post-randomisation); exposure to E-MOTIVE; previous C-section; gestational age; antepartum haemorrhage; retained placenta; and non-source verified blood loss.

The number of imputations, *n*, will be equal to the proportion of patients with any missing primary outcome data. To ensure inferences are valid, clustering will be allowed for in the imputation process. This will be done using BLIMP 3.0 which allow a random effect for cluster to be included [19].

For the MNAR mechanism, we will assume the prevalence of the primary outcome differs in those patients with a missing outcome compared to those patients with an observed primary outcome. To do this, we will include a parameter in the imputation process that shifts the expected prevalence (of the primary outcome) by a fixed amount. This "*missing parameter*" is a fixed value, and the magnitude and sign of the coefficient determines the strength and direction of the missing not at random process (see below). This parameter is included in each arm separately, to allow the MNAR mechanism to influence one arm at a time.

By selecting a variety of values of the missing parameter, we can identify the "tipping point". This refers to the point in which general conclusion of the primary analysis changes. The tipping point will be determined in one of two ways:

Option A: The CI in the primary analysis does not contain one for the risk ratio (i.e., one treatment is superior to the other), then the point in which the CI crosses one will be highlighted (the tipping point). The base case from the primary analysis will be highlighted.

Option B: The CI in the primary analysis does contains one for the risk ratio (i.e., inconclusive whether one treatment is superior to the other), then the point in which the CI no longer contains one will be highlighted (the tipping point). The base case from the primary analysis will be highlighted.

We will consider two scenarios: MNAR in exposed arm only (i.e., those in E-MOTIVE arm in the post-randomisation period) and MNAR in non-exposed arm only (i.e., those in standard care arm and those in baseline period).

**MNAR in exposed arm only**

Firstly, we assume that all patients in the unexposed group with missing outcome data have the same prevalence as those who are observed. In the exposed group, we will set the *missing parameter* to change the prevalence of the outcome in those with missing outcome data in the exposed group. We will estimate the effect measure using *n* imputation, by fitting a model with exposure, time period and cluster-level covariates to each imputed dataset separately. The RR and RD will be obtained in the same manner as for the primary analysis – using marginal standardisation. The summary measure from each of the *n* imputed datasets is then pooled using Rubin's rules. The effect measure and CI will be examined and stored. We will use several values of the *missing parameter* to visualise how the effect measures and confidence intervals change as the prevalence of the outcome in those with missing data changes.

**MNAR in unexposed arm only**

Firstly, we assume that all patients in the exposed group with missing outcome data have the same prevalence as those who are observed. In the unexposed group, we will set the *missing parameter* to change the prevalence of the outcome in those with missing outcome data in the unexposed group. We will estimate the effect measure using *n* imputation, by fitting a model with exposure, time period and cluster-level covariates to each imputed dataset separately. The RR and RD will be obtained in the same manner as for the primary analysis – using marginal standardisation. The summary measure from each of the *n* imputed datasets is then pooled using Rubin's rules. The effect measure and CI will be examined and stored. We will use several values of the *missing parameter* to visualise how the effect measures and confidence intervals change as the prevalence of the outcome in those with missing data changes.

**Independent estimating equations**

Informative cluster sizes may impact the estimate of the effect measure. In the primary analysis methods, we have included a country-centred binary covariate to indicate cluster size. An alternative approach would be to have fitted independent estimating equations – though this reduces power. Therefore, as a sensitivity analysis, we will fit a generalised estimating equation with independent working correlation. We will use log links and identity links to obtain estimates of the risk ratio and risk difference. This is to test whether our approach is robust to informative cluster sizes. This sensitivity analysis was recommended by the statistician from the Data Monitoring Committee.

## 8. Statistical software

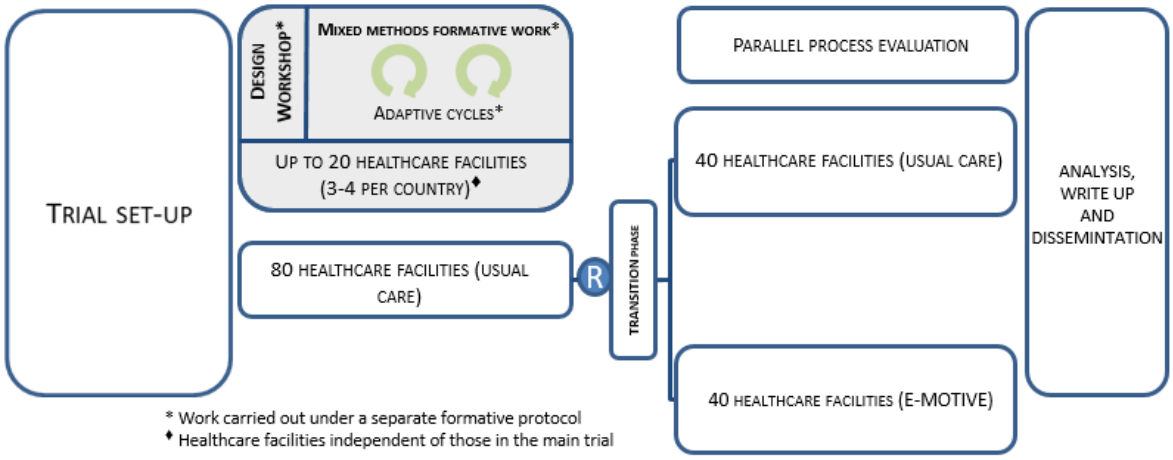
Statistical analysis will be undertaken in the following statistical software packages: Stata.

## Appendix A: Deviations from SAP

This report below follows the statistical analysis plan version *1.0* dated *13/03/2023* apart from following:

| Section of report not following SAP | Reason |
|---|---|
| | |

## Appendix B: Trial schema




## Appendix C: Schedule of assessments

*See trial protocol.*

## Appendix D: Data manipulations

The Trial Statistician will derive all responses from the raw data recorded in the database.

The following variables will be obtained directly from the raw data – where they are recorded in an appropriate form for analysis:

**Cluster demographics**

- Country (Kenya, Nigeria, South Africa, Tanzania)
- Number of births (total, caesarean, and vaginal) during study from weekly facility forms
- Number of births contributing to the study from pregnancy outcome forms
- Number of staff (Consultant Obstetrician; Medical Doctor; Clinical Officer; Medical Student; Nurse; Midwife; Nurse Midwife; Nurse/Midwifery student; other healthcare professional) – data is collected monthly, and the average number of staff in each category per month will be collated for each cluster

**Individual patient demographics**

- Previous births (Numerical)

- Previous caesarean section (Yes or No)
- PPH in previous pregnancy (Yes or No)
- Age (Numerical)
- Index birth
- Multiple pregnancy
- Mode of birth (Spontaneous vaginal or forceps or ventouse)
- Birth weight (Numerical)
- Gestational age at birth (Numerical)
- Antepartum haemorrhage (Yes or No)
- Preeclampsia (Yes or No)
- Labour augmented or induced (Yes or No)
- Retained placenta or manual removal of the placenta (Yes or No)

**Primary Outcome Components**

- Laparotomy for bleeding (Yes or No)
- Maternal mortality for bleeding (Yes or No)

**Secondary/Exploratory Clinical Outcomes**

- Laparotomy (Yes or No)
- Laparotomy with compression sutures (Yes or No)
- Laparotomy with arterial ligation (Yes or No)
- Hysterectomy (Yes or No)
- Hysterectomy for bleeding (Yes or No)
- All cause maternal mortality (Yes or No)
- All cause neonatal mortality (Yes or No)
- Non-pneumatic anti-shock garment (NASG) (Yes or No)
- Uterine balloon tamponade (Yes or No)
- Blood transfusion (Yes or No)
- Blood transfusion for bleeding (Yes or No)
- ICU admissions (Yes or No)
- Transfer to higher level cluster (Yes or No)

**Secondary/Exploratory Implementation Outcomes**

- PPH detection (Yes or No)
- Uterine massage (Yes or No)
- Oxytocin use (Yes or No)
- Misoprostol use (Yes or No)
- Tranexamic acid use (Yes or No)
- Intravenous fluids use (Yes or No)
- Examination of the genital tract (Yes or No)
- Women receiving any treatment uterotonic (Yes or No)
- Women requiring additional treatment interventions (Yes or No)

The following variables will be derived for analysis from the dataset. Details on how they will be derived is presented.

**Cluster demographics**

- Availability of bundle components – A daily facility diary denotes the availability of the bundle components (oxytocin, tranexamic acid, and IV fluid) for each cluster on each day of the study. The proportion of days in which each component is available will be calculated as a ratio of the number of days component was available and the total number of days. The proportion of days in which all components were available will be calculated as a ratio of the number of days with all components available and the total number of days.

**Primary Outcome**

- Primary outcome – The primary outcome is defined as being present if at least one of the three components are present (severe PPH, laparotomy for bleeding, maternal mortality for bleeding).

**Primary Outcome Components**

- Severe PPH – Severe PPH will be defined as present if the blood loss >= 1000ml. The amount of blood loss for each participant will be determined from either:
    1. The weight of the drape (minus 130g which represents the dry weight of the drape for non-calibrated drapes, and minus 120g which represents the dry weight of the calibrated drapes)

**Secondary/Exploratory Clinical Outcomes**

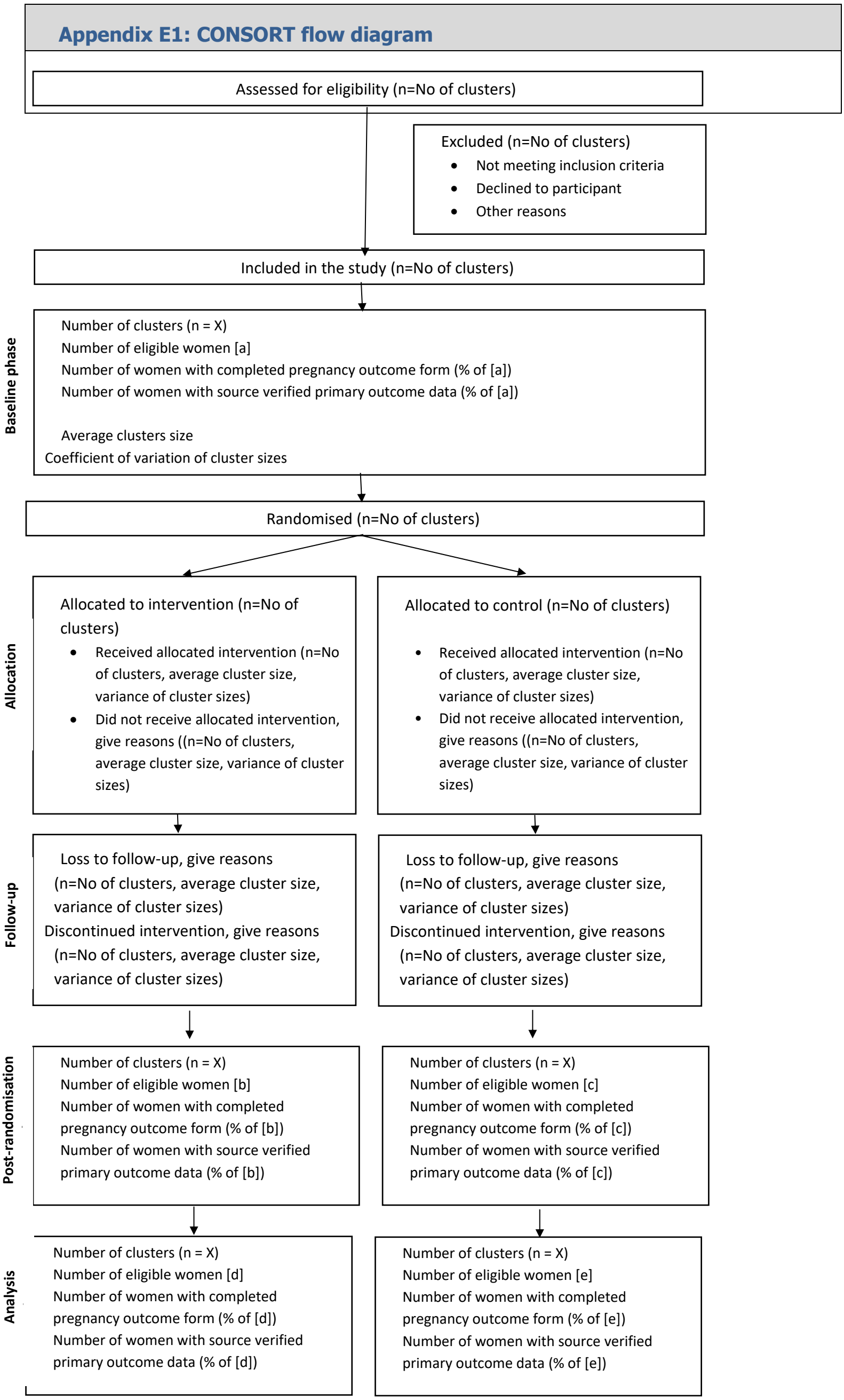
- Primary PPH – defined as present if blood loss is >=500ml. The amount of blood loss is calculated as described above.
- Blood loss 2 hours postpartum - The amount of blood loss for each participant will be determined from either:
    1. The weight of the drape (minus 130g which represents the dry weight of the drape for non-calibrated drapes, and minus 120g which represents the dry weight of the calibrated drapes)

- Blood loss 24 hours postpartum – The amount of blood loss for each participant in the 24 hours will be calculated using the summation of:
    1. Blood loss 2 hours postpartum
    2. Additional blood loss in hours 2 to 24
- Duration of hospitalisation – Calculated as the difference between the date of giving birth and date of discharge.
- Duration of ICU hospitalisation – Calculated as the difference between date of admission to ICU and date of discharge to ICU.

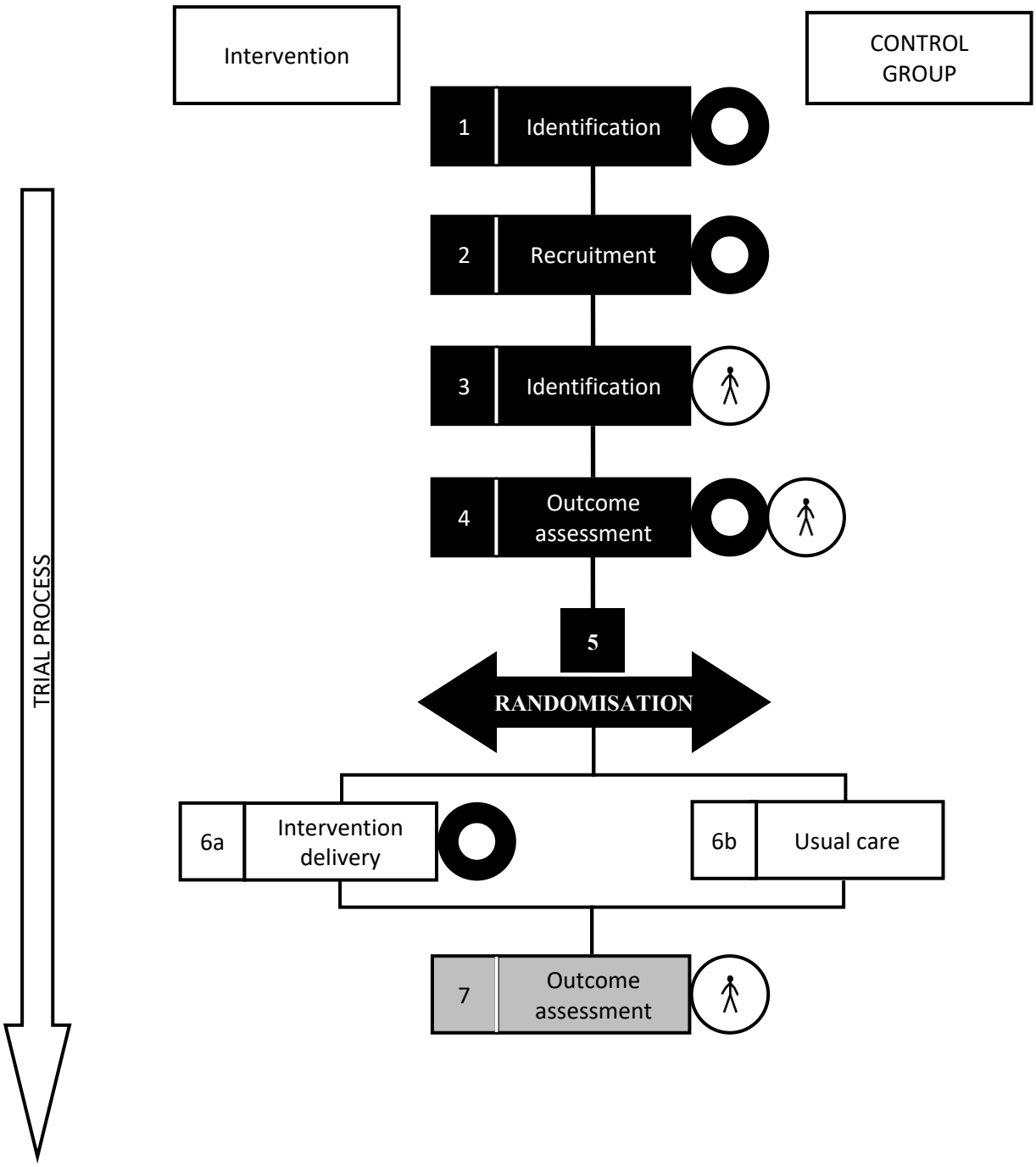
**Secondary Implementation Outcomes**

- Compliance with MOTIVE bundle – Defined as adherence to three core elements of the bundle (oxytocic drugs, tranexamic acid, and IV fluid). Compliance is present if all three elements are present.

## Appendix E1: CONSORT flow diagram



## Appendix E2: Timeline Cluster Graphical Tool Example



| | |
|---|---|
| 1 | Cluster (Health Facility) identification |
| 2 | Cluster recruitment |
| 3 | Patient identification |
| 4 | Outcome assessment – baseline period |
| 5 | Randomisation |
| 6a | Intervention delivery |
| 6b | Usual care |
| 7 | Outcome assessment – post-randomisation period |

## Appendix E3: Baseline characteristics

| Health facility level characteristics | E-MOTIVE | | Control | |
|---|---|---|---|---|
| | Baseline period (N = ) | Post-randomisation period (N = ) | Baseline period (N = ) | Post-randomisation period (N = ) |
| **Facility location** | | | | |
| Country | | | | |
| Kenya | | | | |
| Tanzania | | | | |
| Nigeria | | | | |
| South Africa | | | | |
| **Facility demographics** | | | | |
| Total number of births per facility[2] | | | | |
| Mean (SD)/Median [IQR] | | | | |
| Number of vaginal births per facility[3] | | | | |
| Mean (SD)/Median [IQR] | | | | |
| Number of caesarean sections per facility [4] | | | | |
| Mean (SD)/Median [IQR] | | | | |
| Number of skilled birth attendants | | | | |
| Mean (SD)/Median [IQR] | | | | |

| |
|---|
| Availability of bundle components [6] (months/proportion/??)<br>Mean (SD)/Median [IQR] |

| Patient demographics | | E-MOTIVE | | Control | |
|---|---|---|---|---|---|
| | | Baseline period (N = ) | Post-randomisation period (N = ) | Baseline period (N = ) | Post-randomisation period (N = ) |
| **Previous birth information** | | | | | |
| Previous births | | | | | |
| | 0 | | | | |
| | 1 | | | | |
| | 2 | | | | |
| | 3 | | | | |
| | 4 | | | | |
| | 5 or greater | | | | |
| | Mean (SD)/Median [IQR] | | | | |
| | Missing | | | | |
| Previous caesarean section | | | | | |
| | Missing | | | | |
| PPH in previous pregnancy | | | | | |
| | Missing | | | | |
| Age (years) | | | | | |
| | Mean (SD)/Median [IQR] | | | | |
| | Range | | | | |
| | Missing | | | | |
| **Index birth** | | | | | |
| Multiple pregnancy | | | | | |
| | Singleton | | | | |
| | Multiple pregnancies | | | | |
| | Missing | | | | |
| Mode of birth | | | | | |
| | Spontaneous birth | | | | |
| | Instrumental births | | | | |
| | Missing | | | | |

| | |
|---|---|
| Birth weight (g) | |
| | Mean (SD)/Median [IQR] |
| | Range |
| | Missing |
| Gestational age at birth (weeks) | |
| | Mean (SD)/Median [IQR] |
| | Range |
| | Missing |
| Gestational age <37 weeks | |
| | Missing |
| Antepartum haemorrhage[1] | |
| | Missing |
| Preeclampsia [2] | |
| | Missing |
| Labour augmented or induced | |
| | Missing |
| Retained placenta or manual removal of the placenta[3] | |
| | Missing |

## Appendix E4: Primary outcome results

**Table 3: Main outcomes. Values are number (percentage) unless otherwise stated.**

| Outcomes | E-MOTIVE | | Control | | Risk Ratio (95% CI) | p-value | Risk Difference (95% CI)* | p-value |
|---|---|---|---|---|---|---|---|---|
| | **Baseline period (N = )** | **Post-randomisation (N = )** | **Baseline period (N = )** | **Post-randomisationperiod (N = )** | | | | |
| **Primary Outcome** | | | | | | | | |
| Primary Outcome [2] | | | | | | | | |
| **Key Implementation Outcomes** | | | | | | | | |
| PPH detection [3] | | | | | | - | | - |
| Compliance with MOTIVE bundle [4] | | | | | | - | | - |
| | | | | | | - | | - |

PPH: Post partum haemorrhage. ICU: Intensive care unit.

* Differences between risks are presented in percentage points, and differences between mean values are presented in the unit of the mean values

1: Adjusted for the period in which the observation was made and cluster-level covariates used in the randomisation (number of vaginal births, PPH rate, country, and the primary outcome rate)

2: Primary outcome was a composite of: severe postpartum hemorrhage (blood loss>=1000ml); laporotomy for bleeding; and maternal mortality from bleeding.

3: Defined as recording of diagnosis of PPH by birth attendant.

4: Defined as adherence with three core elements of the bundle: administration of oxytocic drugs, TXA, and IV fluids.

## Appendix E5: Secondary outcomes results

**Table 4: Main outcomes. Values are number (percentage) unless otherwise stated.**

| Outcomes | E-MOTIVE | | Control | | Risk Ratio[1] (95% CI) | Risk Difference[1] (95% CI)* |
|---|---|---|---|---|---|---|
| | Baseline period (N = ) | Post-randomisationperiod (N = ) | Baseline period (N = ) | Post-randomisationperiod (N = ) | | |
| **Secondary Clinical Outcomes** | | | | | | |
| Severe PPH [2] | | | | | | |
| Laparotomy for bleeding | | | | | | |
| Maternal mortality from bleeding | | | | | | |
| Primary PPH [3] | | | | | | |
| Laparotomy | | | | | | |
| Laparotomy with compression sutures | | | | | | |
| Laparotomy with arterial ligation | | | | | | |
| Hysterectomy | | | | | | |
| Hysterectomy for bleeding | | | | | | |
| All cause maternal mortality | | | | | | |
| All cause neonatal mortality | | | | | | |
| Non-pneumatic anti-shock garment (NASG) | | | | | | |
| Uterine balloon tamponade | | | | | | |
| Blood transfusion | | | | | | |
| Blood transfusion for bleeding | | | | | | |
| ICU admissions | | | | | | |
| Transfer to higher level facility | | | | | | |

| | E-MOTIVE | | Control | | | |
|---|---|---|---|---|---|---|
| | **Baseline period (N = )** | **Post-randomisation period (N = )** | **Baseline period (N = )** | **Post-randomisationperiod (N = )** | | **Mean difference[1] (95% CI)*** |
| Blood loss 2 hours postpartum (ml), mean (SD) | | | | | - | |
| Blood loss 24 hours postpartum (ml), mean (SD) | | | | | - | |
| Duration of hospitalisation (hours), mean (SD) | | | | | - | |
| Duration of ICU hospitalisation (hours), mean (SD) | | | | | - | |
| **Other Secondary Implementation Outcomes** | **E-MOTIVE** | | **Control** | | | |
| | **Baseline period (N = )** | **Post-randomisationperiod (N = )** | **Baseline period (N = )** | **Post-randomisationperiod (N = )** | **Risk Ratio[1] (95% CI)** | **Risk Difference[1] (95% CI)*** |
| **In women with ≥500ml blood loss** | | | | | | |
| Uterine massage | | | | | | |
| Oxytocin use | | | | | | |
| Misoprostol use | | | | | | |
| Tranexamic acid use | | | | | | |
| Intravenous fluids use | | | | | | |
| Examination of the genital tract | | | | | | |
| Women receiving any treatment uterotonic | | | | | | |
| Women requiring additional treatment interventions | | | | | | |
| **In women diagnosed with PPH by HCP [4]** | | | | | | |
| Bundle compliance [5] | | | | | | |
| Uterine massage | | | | | | |
| Oxytocin use | | | | | | |

| | | |
|---|---|---|
| Misoprostol use | | |
| Tranexamic acid use | | |
| Intravenous fluids use | | |
| Examination of the genital tract | | |
| Women receiving any treatment uterotonic | | |
| Women requiring additional treatment interventions | | |
| **In all women** | | |
| Detection of PPH by HCP [4] | | |
| Bundle compliance [5] | | |
| Uterine massage | | |
| Oxytocin use | | |
| Misoprostol use | | |
| Tranexamic acid use | | |
| Intravenous fluids use | | |
| Examination of the genital tract | | |
| Women receiving any treatment uterotonic | | |
| Women requiring additional treatment interventions | | |

PPH: Post partum haemorrhage. ICU: Intensive care unit.
* Differences between risks are presented in percentage points, and differences between mean values are presented in the unit of the mean values
1: Adjusted for the period in which the observation was made and cluster-level covariates used in the randomisation (number of vaginal births, country, and the primary outcome rate)
2: Defined as blood loss >=1000ml.
3: Defined as blood loss >=500ml.
4: Defined as recording of diagnosis of PPH by birth attendant.
5: Defined as adherence with three core elements of the bundle: administration of oxytocic drugs, TXA, and IV fluids.

## Appendix E6: Subgroup analysis for primary outcome

**Example Table for subgroup analysis**

| Subgroup description | E-MOTIVE | | Control | | Adjusted summary measure and 95% CI | p-value for interaction | Ratio of ratios |
|---|---|---|---|---|---|---|---|
| | Baseline period (N = ) | Post-randomisationperiod (N = ) | Baseline period (N = ) | Post-randomisationperiod (N = ) | | | |
| Kenya | | | | | | | |
| Nigeria | | | | | | | |
| … | … | | | … | … | … | |

## Appendix E7: Template report

A template report for the final analyses will be provided in a separate document.

## **Appendix Table E8:** *Estimates of the power with changes to key design parameters.*

| | | | RR = 20% | | | RR = 25% | | | RR = 30% | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | Control Prevalence | | | | | | | | |
| | | | 1.50% | 2.00% | 4.00% | 1.50% | 2.00% | 4.00% | 1.50% | 2.00% | 4.00% |
| CAC = 0.95 | ICC | 0.001 | 86.8 | 94.5 | 99.9 | 97.4 | 99.5 | 99.9 | 99.7 | 99.9 | 99.9 |
| | | 0.02 | 43.2 | 54.4 | 84 | 62 | 74.6 | 96.3 | 78.8 | 89.1 | 99.5 |
| | | 0.05 | 24.8 | 31.6 | 56.1 | 36.8 | 46.7 | 76.4 | 50.6 | 62.7 | 90.3 |
| CAC = 0.97 | | 0.001 | 87.5 | 95 | 99.9 | 97.6 | 99.5 | 99.9 | 99.8 | 99.9 | 99.9 |
| | | 0.02 | 53.9 | 66.3 | 92.5 | 74.1 | 85.5 | 99.1 | 88.8 | 95.7 | 99.9 |
| | | 0.05 | 34.1 | 43.4 | 72.7 | 50.2 | 62.4 | 90.1 | 66.8 | 79.1 | 97.7 |
| CAC = 1.00 | | 0.001 | 89.7 | 96.2 | 99.9 | 98.3 | 99.7 | 99.9 | 99.9 | 99.9 | 99.9 |
| | | 0.02 | 83.7 | 92.1 | 99.9 | 95.9 | 99 | 99.9 | 99.5 | 99.9 | 99.9 |
| | | 0.05 | 83 | 92.6 | 99.9 | 96.2 | 99.1 | 99.9 | 99.4 | 99.9 | 99.9 |

*Note: Power has been calculated assuming 72 health facilities, with an average number of 192 births per month (14-month study), and a coefficient of variation of cluster sizes of 0.5. CAC: Cluster autocorrelation; ICC: Intra cluster correlation. RR = Risk reduction*

## Appendix E9: Estimand Information

| Estimand aspect | Definition for different outcomes and analysis populations | | | | | |
|---|---|---|---|---|---|---|
| | **Primary composite outcome (total population)** | **Key implementation outcomes (target population)** | **Binary secondary clinical and implementation outcomes (total population)** | **Continuous secondary outcomes (total population)** | **Binary secondary implementation outcomes (target population)** | |
| ***Population*** | | | | | | |
| Health Facilities | Health facilities that have 1000 to 5000 births a year and provide comprehensive obstetric care with ability to perform surgery for PPH. | | | | | |
| Research Participant | All healthcare providers attending vaginal births at the study facilities. | | | | | |
| Patient | All women having a verified vaginal birth in the study facilities | | | | | |
| ***Treatment conditions*** | E-MOTIVE vs. standard care | | | | | |
| ***Endpoint*** | Composite of three clinical outcomes (see section 4.5) measured until discharge from the health facility | Classification according to outcome at discharge from the health facility | Outcome on continuous measure at discharge from the health facility | Classification according to outcome at discharge from the health facility | Classification according to outcome at discharge from the health facility | |
| ***Population-level summary measure*** | Relative Risk and Risk Difference | Relative Risk and Risk Difference | Relative Risk and Risk Difference | Mean difference. | Relative Risk and Risk Difference | |
| ***Participant- or cluster-average effect*** | Participant-average. | | | | | |
| ***Marginal or conditional effect*** | Conditional | | | | | |
| ***Handling intercurrent events*** | Cluster-level intervention and implementation. Patients cannot prematurely stop treatment. Additional treatments are captured as secondary implementation outcomes. Any harm is captured as secondary clinical outcomes. Patients are followed up to discharge from health facility (or died whilst in the facility). Primary | | | | | |

| | |
|---|---|
| | outcome data can be missing if blood loss data is not source verified or if a case report form is not complete. A sensitivity analysis will evaluate including missing data in the analysis. |

# References


1. Hemming, K., et al., *Interim data monitoring in cluster randomised trials: practical issues and a case study.* Clinical Trials, 2021. **18**(5): p. 552-561.
2. Pocock, S.J. and R. Simon, *Sequential treatment assignment with balancing for prognostic factors in the controlled clinical trial.* Biometrics, 1975: p. 103-115.
3. Food, D., *Administration, Multiple endpoints in clinical trials: guidance for industry.* College Park, Maryland: Food and Drug Administration, 2017.
4. Widmer, M., et al., *Heat-stable carbetocin versus oxytocin to prevent hemorrhage after vaginal birth.* New England Journal of Medicine, 2018. **379**(8): p. 743-752.
5. Campbell, M.K., et al., *Consort 2010 statement: extension to cluster randomised trials.* Bmj, 2012. **345**.
6. Caille, A., et al., *Timeline cluster: a graphical tool to identify risk of bias in cluster randomised trials.* bmj, 2016. **354**.
7. Altman, D.G. and C. Dore, *Randomisation and baseline comparisons in clinical trials.* The Lancet, 1990. **335**(8682): p. 149-153.
8. Hooper, R., et al., *Analysis of cluster randomised trials with an assessment of outcome at baseline.* bmj, 2018. **360**.
9. Kahan, B.C., et al., *Estimands in cluster-randomized trials: choosing analyses that answer the right question.* International Journal of Epidemiology, 2022.
10. Gallis, J.A. and E.L. Turner, *Relative measures of association for binary outcomes: challenges and recommendations for the global health researcher.* Annals of Global Health, 2019. **85**(1).
11. Kirkwood, B.R., et al., *Effect of the Newhints home-visits intervention on neonatal mortality rate and care practices in Ghana: a cluster randomised controlled trial.* The Lancet, 2013. **381**(9884): p. 2184-2192.
12. StataCorp, *Stata 17 Base Reference Manual: Margins*. 2021: College Station, TX: Stata Press.
13. Leyrat, C., et al., *Cluster randomized trials with a small number of clusters: which analyses should be used?* International journal of epidemiology, 2018. **47**(1): p. 321-331.
14. Thompson, J.A., et al., *Cluster randomised trials with a binary outcome and a small number of clusters: comparison of individual and cluster level analysis method.* BMC Medical Research Methodology, 2022. **22**(1): p. 1-15.
15. Li, P. and D.T. Redden, *Comparing denominator degrees of freedom approximations for the generalized linear mixed model in analyzing binary outcome in small sample cluster-randomized trials.* BMC Medical Research Methodology, 2015. **15**(1): p. 1-12.
16. Leyrat, C., et al., *Propensity score methods for estimating relative risks in cluster randomized trials with low-incidence binary outcomes and selection bias.* Statistics in medicine, 2014. **33**(20): p. 3556-3575.
17. Kahan, B.C., et al., *A comparison of methods to adjust for continuous covariates in the analysis of randomised trials.* BMC medical research methodology, 2016. **16**(1): p. 1-10.
18. Grund, S., O. Lüdtke, and A. Robitzsch, *Multiple imputation of missing data for multilevel models: Simulations and recommendations.* Organizational Research Methods, 2018. **21**(1): p. 111-149.
19. Enders, C.K., H. Du, and B.T. Keller, *A model-based imputation procedure for multilevel regression models with random coefficients, interaction effects, and nonlinear terms.* Psychological methods, 2020. **25**(1): p. 88.