Saving Babies' Lives (SBL) study

Statistical Analysis Plan

Dr Kaajal Patel, Dr Verena Carrara, Prof Claudia Turner, Prof Mavuto Mukaka

Version 1.8. Updated 23.09.23

This SAP will be finalised before database lock and analysis.

Contents

ABBREVIATIONS

CHW	community health worker
CONSORT	Consolidated Standards of Reporting Trials
CRT	cluster-randomised trial
ICC	intracluster correlation coefficient
ITT	intention-to-treat
KAP	knowledge, attitudes, practice
KAPES	knowledge, attitudes, practice, equipment, staffing
NMR	neonatal mortality rate
PLA	participatory learning and action
PMR	perinatal mortality rate
PP	per-protocol
SBL	Saving Babies' Lives
SW-CRT	stepped-wedge cluster-randomised trial
VENMR	very early neonatal mortality rate

INTRODUCTION

BACKGROUND

Neonatal mortality (death within the first 28 days of life) remains unacceptably high in many countries, amounting to 2.5 million deaths a year globally [1]. Most neonatal deaths can be prevented with affordable, available interventions [3,4], however many studies successful at reducing neonatal mortality have failed to realise similar gains at scale [5–7]. Effective implementation and scale-up of interventions designed to tackle neonatal mortality is a global health priority. Multifaceted programmes targeting the the whole neonatal period [11] and the whole neonatal healthcare system [4], with sustainability and scalability built into the design, can provide practical insights to solve this challenge. Cambodia has amongst the highest neonatal mortality rates (NMRs) in South-East Asia Asia, with rural areas particularly affected [8].

The primary objective of this study is the design, implementation, and assessment of the Saving Babies' Lives (SBL) programme, a package of interventions designed to reduce neonatal mortality in rural Cambodia.

This study is a five-year stepped-wedge cluster-randomised trial (SW-CRT) conducted in a rural Cambodian province with an estimated annual delivery rate of 6,615. The study is designed to implement and evaluate the SBL programme, which is the intervention. The SBL programme is an iterative package of neonatal interventions spanning the continuum of care and integrating into the existing health system. The SBL programme comprises two major components: participatory learning and action (PLA) with community health workers (CHWs), and capacity building of primary care facilities involving facility-based mentorship. Standard government service continues in control arms. Data collection covering the whole study area includes surveillance of all pregnancies, verbal and social autopsies, and quality of care surveys. Mixed methods data collection supports iteration of the complex intervention, and facilitates impact, outcome, process and economic evaluation.

More details can be found in the published trial protocol [9] and ClinicalTrials.gov registration: NCT04663620.

STUDY DESIGN RATIONALE

A *cluster*-randomised trial (CRT) design was chosen for reasons of pragmatism and generalisability. The SBL intervention targets primary care facilities including their affiliated healthcare workers and CHWs. The intervention does not directly target patients (ie neonates, pregnant women, mothers or families). Thus, the SBL intervention is applied at the level of the population covered by the primary care facilities (and not on a subset of random patients within a primary care facility). The SBL intervention is thus applied at cluster level, with each cluster including an existing primary care facility (and it's affiliated staff and CHWs).

Ethically, delivery of our intervention to clusters rather than randomised individuals is safe. Our intervention investigates the implementation strategy of existing evidence-based neonatal interventions which have already been proven to have clinical utility and in fact included in government guidance. Thus, our intervention poses minimal risk to the population.

The reason to choose a *stepped-wedge* CRT (SW-CRT) design is primarily ethical. A SW-CRT design allows robust evaluation, whilst reconciling ethical constraints by ensuring the intervention, which is likely to be beneficial, reaches all neonates in the province in the quickest way possible. The intervention will be withheld for the minimal time possible. All primary care facilities will receive the intervention by the end of the study, and thus the entire population will receive the expected benefit from the intervention.

Of note, the SW-CRT design provides our study with some additional advantages over a parallel CRT design. Firstly, the SW is logistically advantageous, as it allows role out of our intervention in stages rather than to half of the clusters at once. The latter would involve staffing and resource challenges. Secondly, implementing over several steps allows the SW design to support iteration of the intervention during the study period. The iterative nature of our intervention is a key concept to ensure we create a replicable intervention with the capacity to adapt to the context. Finally, the SW design is more realistic of how the intervention can be scaled up in the future.

Additionally, stepped-wedge designs can help to increase study power when only a small number of clusters are available because each cluster is exposed to both control and interventions conditions. Hence each cluster acts as it's own control and allows within-cluster comparisons. This reduces the sample size required for a stepped-wedge design to reach the same statistical power that would be needed for a parallel cluster design. This is particularly true with higher ICCs.

SUMMARY

This Statistical Analysis Plan (SAP) is written to be consistent with the new Consolidated Standards of Reporting Trials (CONSORT) extension for SW-CRTs [10]. This statistical analysis plan will guide the Trial Statistician during the statistical analysis of all quantitative outcomes in order to answer the objectives of the study.

This SAP will be finalised before database lock and analysis.

STUDY DESIGN

The study will use a five-year stepped-wedge cluster-randomised trial (SW-CRT) design with two arms: the SBL programme intervention and a control. The intervention will be delivered in three sequential steps to several clusters (sequences) per time-period (step) and delivered to all clusters (the units of randomisation) by the end of the study.

Clusters are based on the existing government health system structure. A cluster is defined as a primary care administrative group, as recognised by the Preah Vihear provincial health department, and includes all primary care facilities, primary care workers, CHWs, villagers and villages in that geographical area.

The study area is the whole of Preah Vihear province, which is divided into 21 clusters. The study is planned to take five years, from January 2018 to December 2022, with an initial control period for all clusters, followed by 3 intervention steps (4-9 clusters per step) until all clusters receive the intervention, and ending with a period when the delivery of the intervention has been completed in all clusters.

THE INTERVENTION

The intervention is the SBL programme, an iterative package of implementation strategies to improve neonatal healthcare across the continuum of care, which is designed to integrate into the existing government-run health system. The SBL programme takes 22 months to implement and comprises two parts, which are delivered simultaneously:

- Participatory learning and action (PLA): the community intervention, comprising PLA with CHWs
- Mentorship training: the primary care facility intervention comprising a mentorship training programme

THE CONTROL

Primary care facilities (and their affiliated CHWs and healthcare workers) in control arms receive no intervention; standard government service continues.

RANDOMISATION UNITS & TIME-PERIODS

Randomisation units

The units of randomisation are the participating clusters. In this trial the intervention is implemented at cluster level, in a total of 21 clusters, which covers the whole province (study area) (figure 1). Clusters are based on the existing government health system structure. A cluster is defined as a primary care administrative group, as recognised by the Preah Vihear provincial health department, and includes all primary care facilities, primary care workers, CHWs, villagers and villages in that geographical area. In practice, several small facilities, serving a small population, work as a team (for example, sharing of staff). In these contexts of close association, facilities have been included in one cluster. Changes in primary care facility status, such as newly opened or upgraded facilities, which result in new clusters, are managed according to government definitions. Any changes in clusters will be reported.

Sequences

Sequences are groups of clusters that receive an intervention at each step. Clusters are randomly assigned into three sequences, with 4-9 clusters per sequence.

Steps

Steps are the time-point at which a sequence, or group of clusters, crossover from the control arm to the intervention arm. At each step, the intervention is simultaneously rolled out to all clusters in a sequence. The intervention will be implemented over three steps, until all 21 clusters receive the intervention (figure 2).

Time-periods

The intervention (the SBL programme) is complex and takes time (two years) to deliver. Anticipated effects, such as change in practice, will take time to be established but are expected to be enduring. We do not expect any quantifiable effect on neonatal health outcomes to be realised immediately after starting the intervention.

Some SW-CRTs include a transition period in the design, to account for the time taken for the intervention to be rolled out and create an effect. Data collected during transition periods are excluded from analyses. For our study a transition period is not feasible as the intervention is both complex and novel, making it difficult to estimate or assume when during the two-year intervention period any quantifiable effect might be seen. Instead, we will examine for when after intervention start any effect is seen (lag effect) in the analyses. Thus, in this trial there are no specific time-periods when outcome data will be excluded from the analyses.

Defining time-periods

By default, each cluster starts the study in the control time-period. To define if each cluster has started the intervention at the assigned time point (the assigned step), as per-protocol (PP), we will consider the date of the first day of the first course or the first day of the first PLA meeting (whichever is earlier). We will consider the cluster as having followed as per-protocol (PP) as long as the date of starting both components falls within three months of the planned start date.

Summary

In summary, each cluster is assigned to one of three sequences, and thus each cluster will transition through two time-periods: control (unexposed or before the start of the intervention) and intervention (during the two-year period of intervention). Hence, a neonate will be a ‘control case’ or an ‘intervention case’ depending on when they are born and where they live.

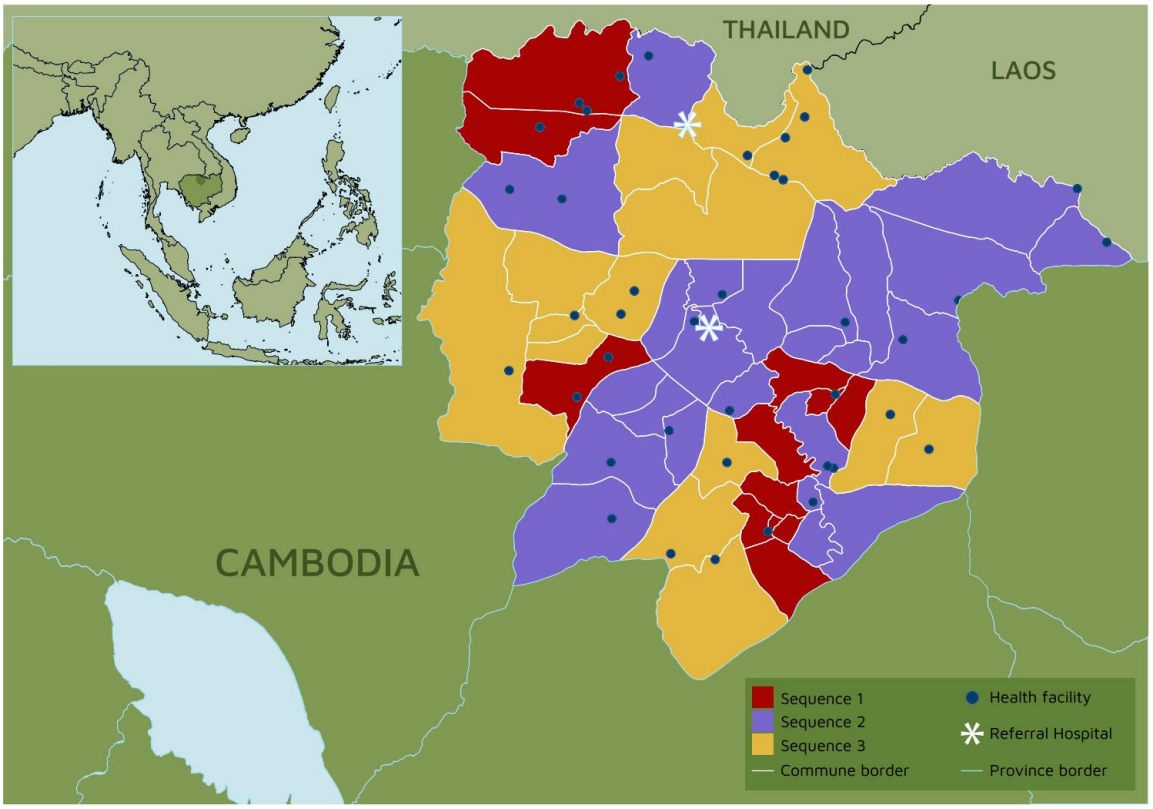

Figure 1: Map of Preah Vihear province, Cambodia showing clusters and health facilities. (own map points; base map data from United Nations Office for the Coordination of Humanitarian Affairs (OCHA), licensed under a Creative Commons Attribution 4.0 International License).

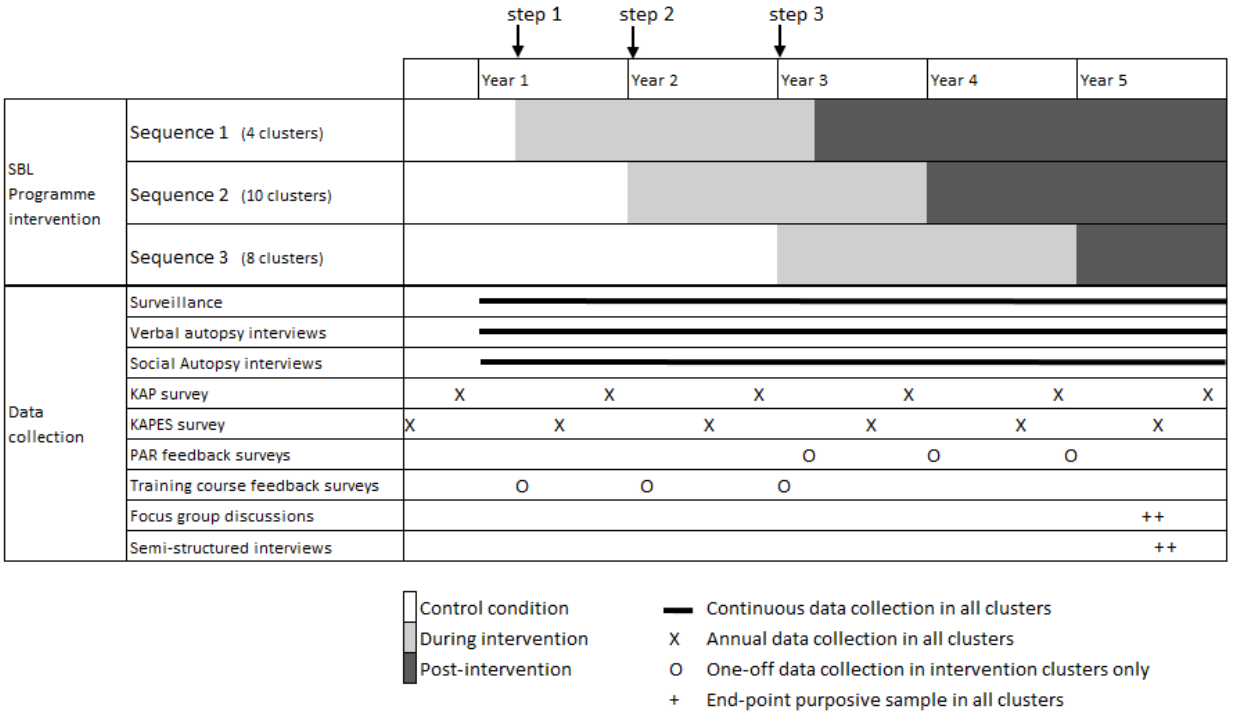

Figure 2: Gantt chart of stepped-wedge study design and data collection.
SBL, Saving Babies' Lives; KAP, knowledge, attitudes, practice; KAPES, knowledge, attitudes, practice, equipment, staffing; PLA, participatory learning and action.

RECRUITMENT

Optimising external validity (or generalisability) is a priority in our study design, as we aim to develop a scalable programme for other provinces in Cambodia. Our trial is also pragmatic and aims to answer whether our intervention can work under real-world conditions and is applicable to a diverse population.

Our trial covers the whole of Preah Vihear province in North-Eastern Cambodia. Our intervention also fully integrates into and strengthens the whole existing provincial health system by the end of the study. By administering our study intervention at the province-level, it will more closely resemble how the scaled intervention could be implemented logistically in the future. By covering the whole province we are also more likely to reach the most vulnerable families living in hard to reach areas. To achieve equity and ensure future programme scale-up reaches marginalised populations where need is often highest, it is important our study covers the whole province without exclusion.

Eligibility of clusters

In our trial, study eligibility has both spatial and individual criteria: area is restricted to geographic clusters, demarcated according to government definitions, participants are all primary care facilities and their affiliated CHWs and healthcare workers working in this area, and the catchment population are all villagers and neonates living in this area.

Eligibility of study participants

All government-run primary care facilities located in the province, including all of their affiliated CHWs and primary care workers, are eligible.

Eligibility of study cases

All deliveries ≥28 weeks gestation occurring in Preah Vihear province or from women living in Preah Vihear province and delivering elsewhere are eligible. All deliveries are followed up until 28 days of neonatal life. Gestation ≥28 weeks will be identified by reported last menstrual period and/or estimated date of delivery when available. When gestational age is unknown, birth weight ≥ one kilogram will be used as a proxy if available. Pregnancy outcomes that are identified as occurring before 28 weeks gestation are not eligible and will be excluded.

RANDOMISATION PROCESS

The study area (Preah Vihear province) is divided into 21 clusters (figure 1), which are subsequently divided into three sequences for cross-over from control to intervention. Appropriate services and authorities in clusters are informed about their crossover as the date approaches.

The first sequence has only four clusters, to incorporate a pilot phase into study design. The four pilot clusters were chosen by the provincial health director, but based on some suggestions of ours to ensure two main factors (convenience sampling):

- Variation of number of health posts associated to the main primary care facility - at least two of the four pilot clusters to have health posts and at least one to have none
- Variation of distance from primary care facility to secondary care (referral) facility

The remaining 17 clusters were pre-assigned to one of the remaining two sequences. As our study had a smaller number of clusters available for randomisation, baseline cluster imbalance was a risk and potentially difficult to avoid with simple randomisation, and therefore we used covariate constrained allocation methodology. Covariate constrained randomisation was performed using r on 3rd February 2019 to balance five cluster-level characteristics:

- Number of villages
- Population size
- Distance from primary care facility to secondary care (referral) facility
- Dry season travel time (mean travel time for each village to primary care facility)
- Wet season travel time (mean travel time for each village to primary care facility)

OUTCOME MEASURES

The complex and long nature of our intervention roll-out and anticipated effect, necessitates observations to be collected in both control and intervention clusters and at strategic time points before, during and after intervention delivery. Outcome measures include both cluster-level and participant-level qualitative and quantitative data. The outcomes of this study will be measured at several levels, primarily: programme outcomes, patient health outcomes of pregnancies ≥28 weeks gestation, participant health system outcomes of community-level and primary care-level neonatal healthcare. Most outcomes will be measured for the entire length of the study-period for all pregnancies ≥28 weeks gestation and all health facilities. If deemed necessary, data collection maybe extended for up to one year. Measurements over time are an open cohort design, meaning pregnant women entering the province or participants moving into intervention clusters will be included.

All objectives, expected outcomes, outcome measures, methods of data collection and time point of data collection are summarised in table 1 in our published manuscript [9].

Our study is innovative as the primary outcome is the design, implementation and assessment of an effective programme to reduce neonatal mortality. This was decided because of our determination to ensure that this study results in an immediately replicable programme should it be proven to be 'successful'. This is also why external validity (or generalisability) and pragmatism are emphasised in our study design.

For our study, both repeated cross-sectional data and longitudinal data are collected. This study is not a longitudinal study of cases, but a study on repeated cross-sectional data on pregnancy outcomes (ie NMRs) occurring in the province.

Pregnancy outcome data is cross-sectional as data in each step in each cluster cannot be the same. Measurements are at the baby level (not maternal level), and each baby can only have one outcome (stillbirth, neonatal death, neonatal survival). A baby is defined as ≥28 weeks gestation and each is followed up until 28 days of completed life. An individual case can only exist (be delivered) once and be followed up once for a maximum of 28 days. Thus, each case can be included only once in the study and multiple outcomes per case cannot occur. Of note, a pregnant woman may deliver more than once during the study period. One pregnancy may result in twins, which would be recorded as two cases. One woman may become pregnant more than once during the five-year study period, in which case each baby would be recorded as a separate case. Thus, each outcome included in the study is distinct. Accounting for

within patient correlation in the analyses is not required.

The NMR per year per cluster will provide data outcomes that can be analysed over a period of time at cluster level (within cluster), as well as comparing intervention and control clusters (between clusters). Change in NMR over time will allow examination of any intervention lag effect after intervention start.

Longitudinal data also include annual quality of care (knowledge, attitudes, practice (KAP) and knowledge, attitudes, practice, equipment, staffing (KAPES)) surveys, which provide repeated data, collected from the same cohort of facilities, which are considered to be the participants. Combining the facility KAPES data per cluster will allow for examination of change over time by cluster (within cluster), as well as investigating differences in results between intervention and control clusters (between clusters). The repeated nature of these measurements will be taken into account when investigating for change in quality of care after intervention start.

DATA SOURCE & COLLECTION

Surveillance

When available routinely collected facility and government administrative data are unreliable. Therefore, to accurately determine the NMR in Preah Vihear a neonatal surveillance system was established. CHWs collect data from their village on all pregnancies ≥28 weeks gestation. All pregnancies are followed up by CHWs until 28 days of life. Data on stillbirths (fetal death ≥28 weeks gestation) and neonatal survival to 28 days are collected.

In addition to outcome data, delivery data on birth weight, gestation, gender and place of birth are also collected, as well as place of death for neonatal deaths. Data are entered into a mobile application by the study team at monthly CHW meetings held in each cluster. Thus, data on delivery outcome and basic characteristics for all births and deaths in the province for the study duration will be collected. No maternal outcomes will be used in analysis.

Verbal & Social autopsy

To determine the medical and social causes of stillbirths and neonatal deaths, a verbal and social autopsy is performed by the study team within six months of death, using an adapted WHO verbal autopsy tool [33,34]. Verbal autopsies are analysed by dual physician analysis and cause of death assigned using the new WHO classification system of ICD-10 for deaths during the perinatal period (ICD-PM) [35,36]. Characteristics of stillbirths and neonatal deaths will also be discerned from verbal autopsy data. Social autopsies will collect data on contributory factors to stillbirths and neonatal deaths, such as social and demographic factors, and barriers and delays to seeking healthcare, structured according to the three-delays framework [37–40]. Thus, data on causes of deaths for all deaths in the province for the study duration will be collected.

KAP & KAPES

We developed a knowledge, attitudes and practice (KAP) survey related to neonatal healthcare for all CHWs in the province. The KAP survey method provides a standardised and comprehensive measure of quality [42]. The KAP survey is conducted annually with CHWs entering answers directly into a mobile application. Thus, data on quality of care for all CHWs in the province for the study duration will be collected.

Primary care facility quality of neonatal healthcare is measured annually using a quantitative and qualitative survey, similar to the CHW KAP survey. With the addition of two further domains to KAP, Equipment and Staffing, we developed a “KAPES” model of assessment for primary care facilities, based on national neonatal guidelines. All primary care workers in the province answer the knowledge and attitudes questions directly into the mobile application. The practice, equipment and staffing components are answered by the study team during annual visits to every primary care facility. Thus, data on quality of care for all primary care facilities in the province for the study duration will be collected.

Cost

A detailed cost analysis will be carried out to estimate additional resources that are needed for the SBL programme over those required for standard care. These costs will be combined with estimates for the incremental cost of caring for neonates at facilities if attendance is found to increase, and by modelling subsequent survival benefits in terms of incremental disability-adjusted life years averted and quality-adjusted life years gained.

POTENTIAL PROBLEMS

As this is an effectiveness study based in real-world conditions over five years, and using a cluster trial design, there are several risks and limitations to be aware of, to examine and to account for in analysis as appropriate [20,21].

MISSING DATA RISKS

Missing pregnancies. By involving CHWs in pregnancy outcome data collection we hope to capture as many cases as possible as they are trusted members of their communities. We hope this is especially supported by the fact that CHWs live in the villages they serve, and so tend to know their fellow villagers. Additionally most pregnancies ≥28 weeks gestation are visible and so we hope any cases that CHWs may not know about might be noticed.

Missing deaths. As the NMR involves a small numerator (neonatal deaths) over a much larger denominator (live births), the significance of missing deaths is important in this study. All primary care facility and hospital records in the province will be checked regularly for stillbirths and neonatal deaths to account for deaths potentially missed by the study surveillance system. This will also help validate death data collected by CHWs.

Missing causes of death. Verbal and social autopsy data collection is particularly sensitive, requiring an interview with families of the deceased. We hope for a refusal rate of less than 25%. Reasons for non-participation in the autopsies will be recorded and we will examine for possible selection bias in our cause of death results with any available data from the surveillance system.

Missing data. Where possible, app-based data collection, such as the KAP and KAPES surveys, will use compulsory answer responses to minimise the potential for missing data collection. The neonatal surveillance system will be checked at the time of entry by the study team, and any missing data discussed with CHWs. However, it is possible that some data will not be known by families or CHWs, such as birth weight if the baby was born at home.

Handling missing data and lost to follow ups

Missing data will be handled using multiple imputation techniques or other comparable robust statistical methods.

STUDY DESIGN RISKS

Cluster imbalance. With CRT trials, each cluster covers a different population, geographical area and contains a different health system-mix. Thus, cluster baseline covariates and cluster size can vary. The resulting confounding of baseline covariates may result in inflated risks of false-positive and false-negative findings, especially with a smaller number of clusters.

Furthermore, with SW-CRT whilst each cluster contributes both control and exposed cases, they will do so in different proportions, depending on when they are randomised to the intervention. This will contribute greatly to any difference between the control and the exposed groups.

We used a covariate constrained randomisation strategy to attempt to avoid cluster imbalance of baseline

covariates and cluster size variability. In case of cluster imbalance in our study we will adjust for baseline characteristics in our analysis.

Risk of clustering. Clustering relates to the interplay between individuals and clusters. Data from individual clusters are not fully independent from each other because of certain selection factors existing in each cluster. An example of clustered data that we may find in our study is families, pregnant women and neonates living in a certain cluster are likely to share similarities such as geography and socioeconomic status. These similarities may have an effect on incidence of home deliveries or neonatal disease burden, for example. Similarly, participants (health workers) in a certain cluster are likely to share similarities. This correlation might lead to an increase of within-cluster correlation and between-cluster variability in neonatal mortality and other study outcomes.

Homogeneity of potential confounders between subjects in clusters can be expected to reduce variability of responses compared to a truly random sample, with a resulting loss of power to detect difference between intervention and control groups. Failing to account for the effects of clustering in the analysis will result in inflated type I error rates. The intracluster correlation coefficient (ICC) is a measure of the relatedness, or similarity, of clustered data. The ICC compares the variance *within* clusters with the variance *between* clusters and can be used to adjust for effects of clustering.

Design effect is a statistical measure expressing the effects of clustering on the sample size (ie. A function of ICC). A cluster trial with a design effect of three requires triple the sample size of an individual RCT. The design effect was factored into the power calculations via the ICC. Furthermore, in our study the clusters are fixed by the study area (province) and the study duration (five years) and therefore gives challenges to apply the design effect accurately to calculate sample size. Thus, the focus was on power calculations based on the fixed number of clusters.

Risk of contamination. All trials involving clusters carry the risk of intervention contamination from intervention clusters into control clusters waiting for crossover. This can reduce the observed effect size as the control clusters cannot be kept as true controls. Clusters based on established groups may reduce the risk of contamination, and so in our study clusters are based on the existing government health system structure. Nevertheless, as this is a complex, long-term intervention set in real-world conditions, some contamination is expected. In our study, contamination risk can be considered in two ways: direct contamination via participants and indirect contamination via cases.

Firstly, direct contamination of the intervention via participants (ie CHWs and primary healthcare staff) may occur by participant movement between clusters. CHWs and primary healthcare workers can move workplace freely during the course of the study. CHWs live in the village they serve. If a CHW moves to live in another cluster, they may also become a CHW in this new cluster. Primary healthcare workers may move to work in another primary care facility, especially since they do not necessarily have to live near their workplace.

It is also possible that participants in control clusters hear about the intervention in the intervention clusters and change their practice accordingly. CHWs mostly interact with other CHWs and other health workers in

their own cluster, as they report to their local health facility and they are therefore unlikely to interact with participants in other clusters. In contrast, all primary facility leaders interact at monthly provincial health meetings and this may especially influence practice by leaders in control facilities as we expect our intervention to bring neonatal healthcare further up the agenda at these monthly meetings. Additionally primary healthcare workers from different facilities quite often attend training courses together and this is another time when participants from different clusters may interact.

Secondly, indirect contamination via families, pregnant women and neonates may occur if they interact with the health system in a different cluster from which they live. Families can move freely across cluster boundaries to receive healthcare. CHWs serve only one village and each village belongs to only one cluster (villages do not span cluster boundaries). Families are therefore likely to only interact with the CHWs in their own village, and the risk of contamination of the community-level intervention is limited. However, risk of contamination of the primary care intervention is greater because most families will need to travel outside of their village to attend a primary care facility. They may therefore travel across a cluster boundary to a primary care facility in another cluster. For example, a pregnant woman may live in a control cluster and so receive no community intervention from the CHW in her village, but for delivery she may choose to travel to a primary care facility in another cluster, if it is located nearer to her home, for example.

Contamination is a limitation of our study design and we will assess this with available data and report accordingly. Observations collected under the control condition will be assessed for contamination by the intervention, and we will account for contamination effects in analysis.

Risk of population movement. Similar to contamination risk, population movement during the study is a challenge. Families may move to live in another cluster. Each case has village of residency recorded, and this is the cluster the case is assigned to. The analysis strategy for NMR will primarily be according to the intention-to-treat (ITT) principle. In addition, per-protocol analyses will be performed as a sensitivity analysis to assess the assumptions made under the ITT principle.

Cluster changes. Our study is set in real world conditions and the government health system around which we have based our cluster geographic definitions may change. For example a new primary care facility may open, which would create a new cluster with surrounding villages (and their families) reassigned to this new cluster. Cluster changes will be managed as they arise, as it will depend on when they occur in the study (if it is in the control, intervention or post-intervention group). We are committed to ensure intervention role out to the whole province by the end of the study and so any new clusters arising after the intervention has started in sequence two, will be assigned to sequence three by default. All cluster changes will be reported and accounted for during analysis and explored as 'per-protocol'. Analysis of 'as-treated' might also be considered in secondary analysis if appropriate.

Cluster non-blinding. For our study it is impossible to blind the study team or participants. We will inform clusters of the intervention as it approaches. The provincial health government has informed all facility leaders that the intervention will be rolled out to all facilities in the province over five years. After the second sequence, the remaining clusters will be aware that they are next.

Cluster drop out. It is possible that a cluster will drop-out from the trial after being randomised (and so receive no intervention at all). This is very unlikely as the provincial government regards participation of primary care

facilities, CHWs and healthcare workers with our intervention as being part of their normal role. However, if this occurs, the impact will be explored in a per-protocol analysis.

Risk of following timeline as 'per-protocol'. Logistically this is a challenging intervention to implement in rural Cambodia and there is a risk of falling behind on timelines and subsequently the crossover of each sequence from control to intervention maybe delayed. Possible reasons for delays include logistical reasons, such as inability to conduct activities due to flooding and dangerous road conditions, and study team staff turnover, especially for the mentorship component part of the primary care intervention for which staff are required to be experienced, with neonatal expertise and to work alone (compared to the PLA component where staff work in teams). The impact of any delays will be explored as 'per-protocol'. Analysis of 'as-treated' might also be considered in secondary analysis if appropriate.

Risk of temporal trends. A particular characteristic of SW-CRT over CRTs is the confounding effect of time. Calendar time may be associated with the outcome in addition to it's association with the intervention exposure, and so time is a potential confounder. Analysis will take into account the confounding effects of time.

SAMPLE SIZE & POWER CALCULATIONS

The recorded NMR for Cambodia was 18.4 per 1000 live births (interquartile range: 10.7 – 28.9) [8]. The specific NMR for Preah Vihear province is not available. The NMR in rural Cambodia is higher than in urban areas [8]. Preah Vihear is predominantly rural, so for the purposes of the sample size calculation an NMR of 28.9 per 1000 live births was used. Preah Vihear province has 6,615 deliveries per year (data obtained from Preah Vihear provincial health records), so an estimated 33,000 deliveries can be expected during the five-year study period.

It is reasonable to propose that the SBL intervention will cause a one-third reduction in the NMR [3,20], from approximately 29 per 1000 to around 19 per 1000 live births. Considering the stepped-wedge design with three steps, total 21 clusters and using an intracluster correlation coefficient (ICC) of 0.05, this study has approximately 80% power to detect a one-third drop in NMR due to the intervention. A total of approximately 26,500 (i.e. 26,460) neonates will be required. A two-sided alpha of 0.05 was used in sample size and power calculations. An ICC of 0.05 used for power calculations in the stepped-wedge design maximizes the sample size [21].

STATISTICAL ANALYSIS

Analysis of SBL programme effectiveness will be based on comparison of intervention and control groups according to the SW-CRT design and will allow for clustering and the confounding effects of time. Covariates used in the constrained randomisation will be adjusted for in the statistical analyses. Descriptive analysis and logistic regression will be used to compare intervention and control groups, and time points. NMR will be described by hazard rate and 95% confidence interval separated by cluster groups. Survival between groups will also be described using Kaplan-Meier plots. Log-rank test will be used for univariate comparison of survival. Epidemiological analysis will be performed on datasets from the surveillance system, verbal autopsy and social autopsy.

All analysis will be performed using the R software package [22] or an alternative software. The main strategy of analysis for the NMR outcome will be according to the intention-to-treat (ITT) principle. Per-protocol analyses will also be performed as a form of sensitivity analysis to the assumption of the ITT approach. Recognised risks of SW-CRTs, including methodological challenges and biases of temporal effects, intra-cluster contamination, and non-blinding of clusters, will be examined and reported [10,23]. The final study report will follow the new Consolidated Standards of Reporting Trials extension for SW-CRTs [10].

No interim analysis will be conducted on the outcomes. However, regular data monitoring will be done to ensure that the data collected by the CHWs and study team are in the right format and complete.

DESCRIPTIVE STATISTICS

Descriptive statistics will be used to present the baseline covariate data, baseline case data, and main outcomes results. We will present categorical variables as numbers and percentages. Continuous variables will be presented using mean and standard-deviation, or median and interquartile range, depending on their distribution.

BASELINE DATA

BASELINE CLUSTER CHARACTERISTICS

The baseline characteristics of clusters will be presented in table 1a. The purpose of the table is to describe the baseline characteristics of each cluster at the start of the baseline and the start of the intervention period to check for any cluster imbalance in cluster characteristics and also to examine if there is any change in cluster characteristics whilst waiting to receive the intervention. Despite our randomisation process to try to balance cluster characteristics, imbalance between clusters and between sequences is possible and if found we will adjust the analyses for them.

The characteristics we will include in the table are: the five parameters used in the covariate constrained randomisation process, the additional parameter used to select the pilot clusters, and a few other characteristics for which we have data and which we consider as a potential confounder.

Summary data will be described per cluster and sequence, using mean and standard deviation, or median and interquartile range, depending on distribution.

Table 1a: Cluster baseline covariates during each cluster's control and intervention time periods.

** covariates included in randomisation procedure. # covariate included in choice of Pilot clusters*

Cluster	Population * n	Number of villages * n	Number of CHWs n	Number of primary healthcare staff n	Number of primary care facilities n	Distance from primary care facility to secondary care (referral) facility * (km)	Dry season travel time (mean travel time for each village to primary care facility) * (minutes)	Wet season travel time (mean travel time for each village to primary care facility) * (minutes)	Rurality (rural or urban)	etc
A Control Intervention										
B Control Intervention etc										
Subtotal Sequence 1 Control Intervention										

BASELINE CASES

We used a covariate constrained randomisation strategy to avoid cluster imbalance of the size of the clusters. Prior to study start accurate cluster-level delivery and outcome data was not available to use as a covariate for the randomisation process. We will therefore need to assess for cluster size imbalance with regards to the number of deliveries per cluster, as well as delivery outcomes.

We will present table 1b with number of deliveries and delivery outcomes (stillbirth, neonatal death, survival and lost to follow up) per cluster during the baseline period (initial three months when all clusters are in the control condition). The purpose of the table is to describe the number of deliveries and delivery outcomes per cluster in order to check for any cluster imbalance.

Table 1b: Baseline cases (number of deliveries and their outcomes) during the initial three month baseline period when all clusters were in the control condition

Cluster	N deliveries	N stillbirths	N neonatal deaths	N neonatal survival	N lost to follow up
A	50	1	2	47	0
B	etc				
Sequence 1 (subtotal)	etc				

OUTCOME DATA

CONSORT flow chart / Trial Profile

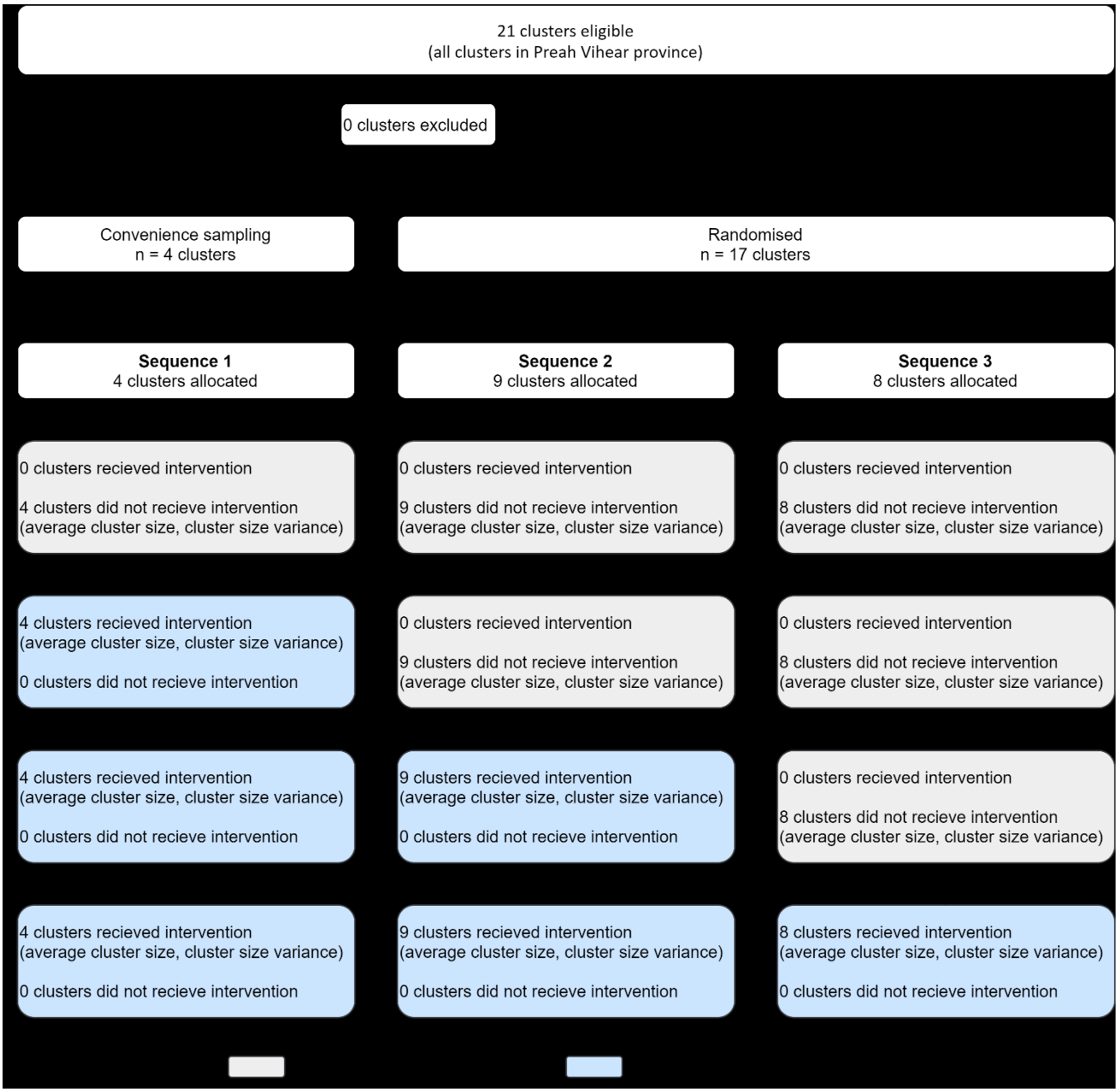

Figure 3: Cluster flowchart for the Saving Babies Lives stepped-wedge cluster-randomised trial by allocated sequence and step. Cluster size = number of deliveries per month

Data completeness

We will present a table with lost to follow up per cluster per time-period (control, intervention), and data completeness of covariates (birth weight, gender, gestation) for neonatal deaths and neonatal survivals. The purpose of the table is to present the outcomes per study time-period and assess for data completeness of the covariates that will be used in secondary outcome analysis (birth weight, gender, gestation).

Table 2: data completeness per cluster by control and intervention time-period.

** Birth weight, gender, gestation*

Cluster	Deliveries n	Neonatal deaths n	Neonatal survival n	Lost to follow up n (% of deliveries)	Neonatal deaths and complete covariates* n (% of neonatal deaths)	Neonatal survivals and complete covariates* n (% of neonatal survivals)
A Control Intervention	50 etc	2	47	1	1 (50%)	30 (64%)
B Control Intervention etc						
Sequence 1 (subtotal) Control Intervention etc						

Delivery outcome data

We will present a table with delivery outcomes and individual patient characteristics per cluster, by intervention condition (control, intervention). The purpose of the table is to present the unadjusted and adjusted secondary outcome results, alongside patient-level characteristics at the time of birth.

Table 3: unadjusted and adjusted delivery outcomes (NMR, SBR) and patient-level characteristics by cluster and intervention condition

		Univariate			Multivariable		
Cluster	Variable	Control	Intervention	P-value	Control	Intervention	P-value
A	**Cases** N deliveries **Outcome** Stillbirths n (SBR per 1000) Neonatal deaths n (NMR per 1000) Neonatal survival n Lost to follow up n **Gender** Male n (%) N missing n (%) **Gestation** Preterm n (%) Missing n (%) **Birth weight** n (median, IQR) Missing n (%) **Place of birth** Community n (%) Primary care facility n (%) Secondary care facility n (%) Other n (%) Missing n (%) **Place of death** Community n (%) Primary care facility n (%) Secondary care facility n (%) Other n (%) Missing n (%) etc						
B	etc						
Sequence 1 (subtotal)	etc						

More detailed graphical representation of the secondary outcome (neonatal mortality) will be given. For example we will present neonatal mortality over time by cluster and by sequence (figure 4). Survival between groups will also be described using Kaplan-Meier plots. Summary results of tertiary outcomes will also be presented in graphical or table format.

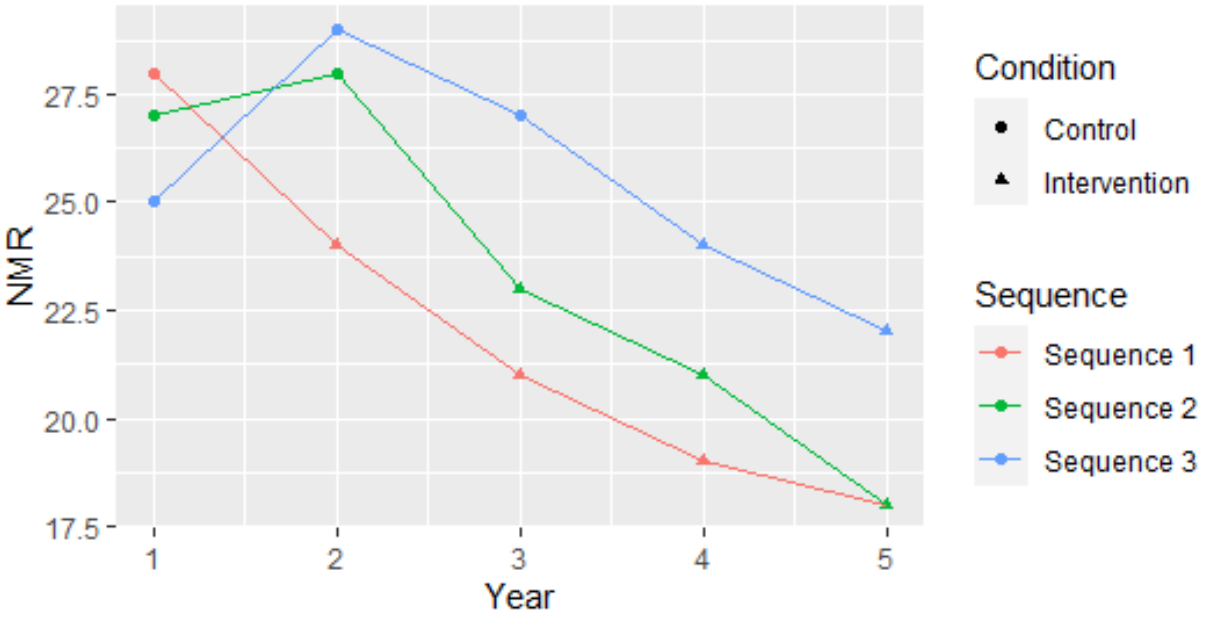

Figure 4: Neonatal mortality rate (per 1000 live births) over time, shown by sequence and study condition

ANALYSIS

ANALYSIS OF PRIMARY OUTCOME

The primary objective of this study is the design, implementation and assessment of an effective programme to reduce neonatal mortality. This will be measured by replication of the SBL programme blueprint into another province (essentially a binary outcome: yes, no). The primary outcome will depend on both the secondary outcome, and to a varying degree the results of the tertiary outcomes. Thus, results from the secondary and tertiary outcomes will support and provide the context to interpret the primary outcome.

ANALYSIS OF SECONDARY OUTCOME

Analysis of the secondary outcome (NMR) will be done using a mixed-effects logistic regression, as a patient-level analysis. The outcome response will be binary (neonatal death, neonatal survival). NMR, calculated as number of neonatal deaths before 28 days per 1000 live births ≥28 weeks gestation, is expected to decrease after the intervention.

The primary analysis of the secondary outcome (NMR) will be performed on an intention-to-treat basis (ITT), with all clusters in a sequence considered to have followed the protocol if the first day of the first course or the first day of the first PLA meeting (whichever is earlier) started within the three months the sequence was allocated to. As this is a complex intervention to roll out, some deviation from protocol is expected to occur. In this case a per-protocol (PP) analysis will be performed. An as-treated analysis may also be performed as appropriate. Results will be presented as hazard rate with 95% CI.

Modelling

We will use the Hussey and Hughes model (26). This model makes two important assumptions: the underlying secular trend is the same in all clusters, and the treatment produces the same, constant shift in this trend in all clusters and over all time-points. Thus, we will apply appropriate model extensions to adjust for relevant secular trends and treatment-effects.

The hazard rate of the mortality risk for the treatment effect (intervention versus control) with 95% confidence interval will be presented (Model-1). Specific to SW-CRT designs, analysis will be adjusted by time-period (step) (Model-2). Subsequently we will adjust for individual patient characteristics (Model-3) such as birth weight, gender, gestation. We will then explore for possible interaction between time and treatment effect (Model-4).

The results from Model-3 will be considered the primary result, as the aim of this trial was to determine if any change after treatment in short-term mortality is related to the intervention and not to an independent calendar time trend, and since the secondary outcome, neonatal mortality, is highly correlated to patient characteristics, the results adjusted by patient data are believed to be the most appropriate.

Model building

Model-1

We will start with an unadjusted before/after analysis of the effect of intervention (ignoring time effect).

Model-2

Then we'll build in the effect of time-period (step) to investigate if any potential treatment effect is related only to the treatment or also to an independent effect of calendar time. Calendar time could be a potential confounder as other factors/events (for example other changes in government practice) could influence the outcome measure in both control and exposed patients ('rising tide'). As this effect could be either absent or gradual (progressive slow trend), calendar time will be fitted in the model as a linear variable.

Model-3

Then adjustment for patient-level characteristics at time of birth will be included in the model. The covariates of adjustment used in this analysis will be the following: birth weight (linear or binary: in kilograms or low, normal/high as needed), gender (binary: male, female), gestation (binary: preterm, term).

Model-4

Further to this, time will also be fitted as time since intervention start (time as treatment effect modifier), to examine how the impact of the intervention develops over time. The potential treatment effect could increase over time with increased experience of staff, but could also decrease over time after an initial improvement (as enthusiasm decreases, staff turnover, knowledge attrition etc). Therefore, we will explore how long it takes to see a full size effect of the intervention over the secondary outcome (lag effect) and if the size of the effect is maintained or wanes over time (decay effect).

ANALYSIS OF TERTIARY OUTCOMES

Count data will be compared using a multi-level Poisson model with a log link and with an offset term to account for variation due to lengths of time in the control and exposed time periods. We will include multi-level random effects to account for nested clustering within the data and one fixed effect term for each time period (control and exposed). The model will be adjusted for the same covariates as considered for the secondary outcome. Results will be presented as beta estimates with 95% CI.

Binary data will be compared using a multi-level mixed-effects logistic regression models. We will include multi-level random effects to account for nested clustering within the data and one fixed effect term for each time period (control and exposed). The model will be adjusted for the same covariates as considered for the secondary outcome. Results will be presented as relative risk with 95% CI.

Mortality outcomes

All tertiary mortality outcomes (facility NMR, overall perinatal mortality rate (PMR), very early NMR (VENMR)) will undergo the same analysis as for the secondary analysis.

KAP/KAPES

The KAP and KAPES surveys measure quality of care at the primary facility-level. They involve several data collection methods (see separate excel document):

- K (KAP, KAPES): binary data (true/false), participant-level (CHW / healthcare worker)

- A (KAP, KAPES): 5-point likert scale (very unlikely/unlikely/neutral/likely/very likely), collected at participant-level (CHW / healthcare worker)
- P (KAP, KAPES): combination of response options (binary: yes/no, categorical, likert), collected at participant-level (CHW) or primary facility-level (by study team)
- E (KAPES): binary data, collected at primary facility-level (by study team)
- S (KAPES): combination of response options (binary: yes/no, categorical, count), collected at primary facility-level (by study team)

Analysis for quality of care outcomes (results from KAP and KAPES surveys) will be conducted at the primary facility-level per component (ie K/A/P/E/S) and overall (ie combined score). Summary results for each component will also be analysed by four clinically relevant topics: neonatal resuscitation, routine care, recognition of the sick neonate, infection prevention control.

Cost-effectiveness analysis

A detailed cost analysis will be carried out to estimate additional resources that are needed for the SBL programme over those required for standard care. These costs will be combined with estimates for the incremental cost of caring for neonates at facilities if attendance is found to increase, and by modelling subsequent survival benefits in terms of incremental disability-adjusted life years averted and quality-adjusted life years gained. Cost-effectiveness of the SBL programme compared to the normal standard of care (control arm) will be assessed.

ADDENDUM 1: CHANGE TO SECONDARY OUTCOME, JANUARY 2021

Summary: secondary outcome measure changed from neonatal mortality rate to perinatal mortality rate.

RATIONALE

Our original sample size projections and power calculations for the secondary outcome: neonatal mortality rate (NMR) reduction were based on the best data available at the time, as the specific NMR for Preah Vihear province was not available at the start of the study. Data available included an NMR of 18.4 per 1000 live births (25^{th} and 75^{th} centiles: 10.7 – 28.9) for Cambodia in 2012 and an NMR three times higher in rural areas as compared to urban areas [8]. Thus, we used an NMR of 28.9 per 1000 live births for our original power calculations.

However, during our first two years (September 2018 – September 2020) of data collection the NMR in Preah Vihear province was found to be lower than predicted. The perinatal mortality rate (PMR: stillbirths and neonatal deaths within the first seven days of life) is more in line with our original predictions. The PMR is approximately 19.8 per 1000 deliveries in Preah Vihear province from the first two years of data collection.

SAMPLE SIZE & POWER CALCULATION 14 JAN 2021

To calculate the power of PMR as the secondary outcome instead of NMR, we applied a PMR of 19.82 per 1000 deliveries instead of an NMR of 28.9 per 1000 live births to our calculations. We used the same stepped-wedged design assumptions during power calculation for PMR as we did originally for NMR: one-third reduction in the PMR from approximately 19.8 to 13 per 1000 deliveries in year five of the study; 250 deliveries per cluster per year; 3 steps; an average of 7 clusters per step; and in the absence of a known ICC, a reasonably high intracluster-correlation coefficient (ICC) of 0.5. The power and sample size calculations were performed in Stata 16.0 [27].

For the purposes of the power calculation of the secondary outcome, a PMR of 19.82 per 1000 deliveries provides approximately **82% power** to detect a one-third drop in PMR due to the SBL intervention.

From this power calculation a total of approximately 21,000 deliveries will be expected to be studied on average. We are finding ~5800 deliveries / year. Since the three steps will span to a five-year period, 29,000 deliveries will be expected by the end of the five-year period. This higher number of deliveries will increase the power of the study further.

ADDENDUM 2: CHANGE TO STUDY DESIGN, OCTOBER 2021

Summary: study design changed from stepped-wedge cluster-randomised trial to cluster randomised trial.

RATIONALE

A stepped-wedge cluster-randomised trial (SW-CRT) design was chosen over a simple parallel CRT for primarily ethical reasons. We wanted to roll-out the intervention, which we believed to be safe and beneficial, to the whole study area as quickly as possible over the five-year study period, whilst ensuring robust evaluation is possible.

The Covid-19 pandemic began in early year three of the study, whilst the intervention was being rolled out in sequences one and two. The impact of the Covid-19 pandemic on our study has been primarily to the timeline.

National, regional and local level travel restrictions and case outbreaks have meant the study team have been intermittently unable to travel to study areas for data collection and intervention implementation. Additionally, when the team were able to enter study areas, a 'catch-up' period for data collection was needed. Complying with government Covid-19 regulations, such as limiting group size, social distancing, mask wearing, working in outdoor spaces and hand hygiene were strictly followed to protect the study team and participants. Inability to travel to study areas and catch-up periods, as well as reduced group sizes, which placed pressure on staffing, all contributed to the intervention taking longer to roll-out.

Gaps in implementation due to the many study pauses also led to a loss of momentum in both health centre training and in PAR, and a refresher was often needed on return to fieldwork. An additional challenge of the Covid-19 pandemic was that participants were usually busy with Covid-19 work, such as case tracking or vaccination drives. Health centre staff were often absent during mentorship visits, and the staff remaining were stretched with the running of the health centre. A combination of these factors contributed to a dilution effect on the SBL intervention, as implementation could not maintain the planned regularity, intensity or reach.

Altogether, this major change in the context within which we were conducting our study resulted in a significant extension to the timeline (figure 5). As a consequence of the extended time to complete the intervention in sequences one and two, they were still incomplete by near the end of year four of the study (figure 5). Due to human resource constraints we were unable to commence step three until the intervention in step one was completed. As of September 2021, to complete the SW-CRT design a two-year extension to the study was needed (figure 5), which would require considerable extra funding.

Given the dilution effect of the intervention due to the Covid-19 pandemic, we no longer expected to see a significant reduction in PMR due to the SBL intervention. The cost of continuing the study for a further two years, plus possible further extensions, was felt by the principal investigators to be unethical. Instead, we decided to change the study design from a stepped-wedge CRT to a simple, parallel CRT design, with sequence two forming the intervention arm and sequence three forming the control arm (figure 5). We would then roll-out the intervention into control clusters as soon as the study ends. Sequence one included the four pilot clusters, which were not truly randomised and they will be excluded from the main study and primary analysis.

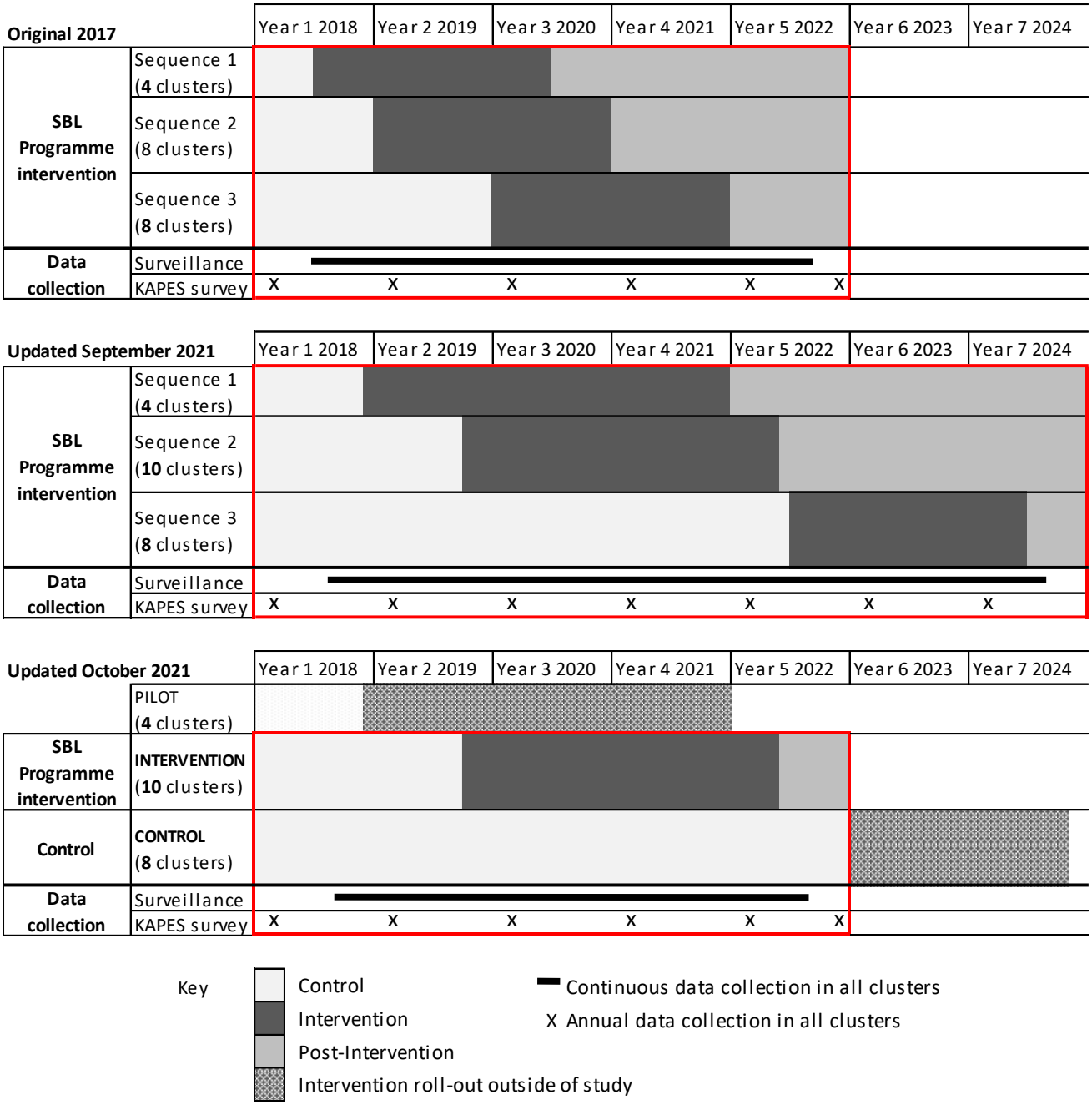

Figure 5: Study timelines over time. Clusters included in the study are contained within the red box in each timeline.

Top timeline: the original stepped-wedge cluster-randomised trial (SW-CRT) design over five-years with three steps.

Middle timeline: the study's SW-CRT design updated in September 2021, 1.5 years into the Covid-19 pandemic. Intervention roll-out took much longer than planned in sequences one and two due to the Covid-19 pandemic. The intervention in sequence three clusters had not yet started. The study would need to be extended by (at least) two years to complete the intervention roll-out in all three sequences as per a SW-CRT study design.

Bottom timeline: the study timeline after changing from a SW-CRT design to a parallel CRT design in October 2021. The former sequence one clusters become the pilot clusters, which are excluded from the main CRT study. The former sequence two clusters become the intervention arm and the former sequence three clusters become the control arm. The study would be able to complete within five years. The control clusters will receive the intervention as soon as the study ends.

IMPLICATIONS TO THE ORIGINAL SW-CRT SAP

Changes and considerations to the SAP that are required based on the change in study design from SW-CRT to a parallel CRT will be covered below systematically by section of the original SAP. The SAP will now follow the Consolidated Standards of Reporting Trials (CONSORT) for CRTs [25] rather than the CONSORT extension for SW-CRTs [10].

STUDY DESIGN

The study will use a cluster-randomised trial (CRT) design with two arms: the SBL programme intervention and a control. Only intervention arm clusters (former sequence two) will receive the SBL intervention.

The study area remains the same except it is smaller by four clusters, since the pilot clusters (former sequence one) will be excluded from the main analysis. The overall study time-period remains the same: five years from January 2018 to December 2022. Initially there is a control period for all clusters (both intervention and control arms), and then the intervention is rolled-out into the intervention arm clusters. By study end, the intervention will have been completed in the intervention arm clusters for a few months.

THE INTERVENTION

The intervention remains the same. The only difference is the originally planned 22-month intervention will take 30 months due to Covid-19 related restrictions and delays to implementation.

THE CONTROL

No change. The control clusters remain as controls throughout the study.

RANDOMISATION UNITS & TIME-PERIODS

Randomisation units

Clusters remain the units of randomisation, and their definitions according to government definitions remain unchanged. Previously conducted randomised assignment of clusters to each sequence remains unchanged.

Sequences no longer exist. Sequences have been turned into parallel randomisation arms: sequence one clusters are now pilot clusters, sequence two clusters are now intervention arm clusters, sequence three clusters are now control arm clusters.

Steps no longer exist.

Time-periods

We are aware that our intervention may take time to have an effect. Therefore, we will examine lag effect of the intervention on perinatal mortality in a subgroup analysis. Specific time-periods will be considered especially for the intervention group.

Defining time-periods

Steps are not relevant for a CRT design. We will consider in the analysis the actual start time of the intervention in each cluster.

Summary

In summary, only the intervention arm clusters will transition from control to intervention conditions. The control clusters will remain in the control condition for the duration of the study.

RECRUITMENT

We will not be able to cover the whole province with our intervention during the study, as the control clusters will not receive any intervention. However, we are committed to rolling out the intervention as a stand-alone programme (without any data collection) as soon as the study ends.

Eligibility of clusters.

No change.

Eligibility of study participants.

No change.

Eligibility of study cases.

No change.

RANDOMISATION PROCESS

No change. Prior randomisation of clusters into sequence two and three will be adhered to. Sequence two clusters are now intervention arm clusters. Sequence three clusters are now control arm clusters.

OUTCOME MEASURES

No change to primary, secondary and tertiary outcomes.

DATA SOURCE & COLLECTION

Surveillance.

No change.

Verbal & Social autopsy.

No change.

KAP & KAPES.

No change.

Cost.

No change.

POTENTIAL PROBLEMS

MISSING DATA RISKS

Missing pregnancies. No change.

Missing deaths. No change.

Missing causes of death. No change.

Missing data. No change.

Handling missing data and lost to follow ups. No change.

STUDY DESIGN RISKS

Cluster imbalance. No change. Of note, with a parallel CRT compared to a SW-CRT, each cluster does not contribute to both control and exposed cases. Only the intervention arm clusters will do this now.

Risk of clustering. No change.

Design effect. No change except to note that the fixed number of clusters is less as it excludes the four pilot clusters.

Risk of contamination. Of note, the risk of intervention contamination from intervention clusters into control clusters will exist for the duration of the study, as the control clusters will not crossover to the intervention during the study. The longer duration may increase the potential contamination effect and reduce the observed effect size as the control clusters cannot be kept as true controls. Contamination remains a major limitation of our study design. Observations collected in the control arm will be assessed for contamination by the intervention, and we will account for contamination effects in analysis.

Risk of population movement. No change.

Cluster changes. No change except to note that any new clusters arising after the intervention has started, will be assigned to the control arm by default. All cluster changes will be reported and accounted for during analysis and explored as 'per-protocol'. Analysis of 'as-treated' might also be considered in secondary analysis if appropriate.

Cluster non-blinding. No change.

Cluster drop out. No change.

Risk of following timeline as 'per-protocol'. With the simplified CRT study design, the risk of delays of crossover of each sequence from control to intervention is no longer a consideration or risk.

Risk of temporal trends. A particular characteristic of SW-CRT over CRTs is the confounding effect of time. With this simplified CRT study design, time will no longer be a potential confounder and analysis will not need to take into account the confounding effects of time. We will adjust for the baseline survey information before the start of the intervention.

SAMPLE SIZE AND POWER CALCULATIONS

Number of clusters and time-frame are fixed to the intervention and control arm clusters and five years, respectively. The study now only excludes the four pilot clusters. The change in study design has been necessitated by logistical reasons and the post-study power calculations are not relevant. We will report limitations to the power in write-up.

STATISTICAL ANALYSIS

The statistical analysis will remain broadly the same, with the main difference being that we will no longer adjust for the confounding effects of time.

The primary analysis will exclude pilot cluster results, and will only be conducted on the intervention and control arm clusters that were randomised at baseline (intention-to-treat approach). We will conduct sub-analyses separately to include the pilot cluster results. Additionally, per-protocol analyses will be done to account for changes in clusters and/or new clusters which were not randomised.

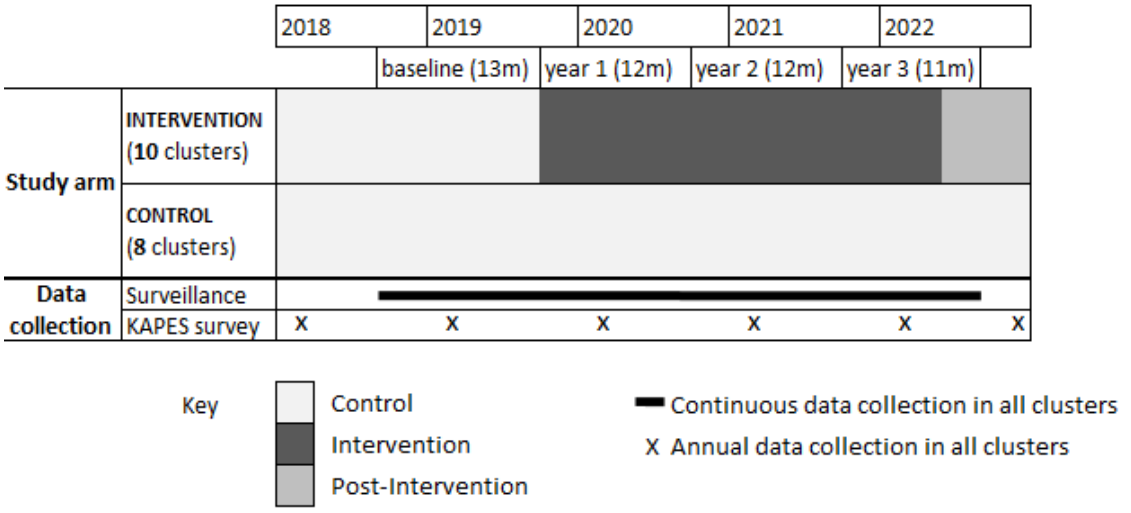

Figure 6: We will present cross-sectional results longitudinally according to four main time periods: baseline, year 1, year 2, year 3

PLANNED ANALYSIS POPULATIONS

Intention-to-treat populations

Clusters which were allocated at baseline will be included in the intention-to-treat populations. These 'baseline' clusters are based on the cluster of residence of each case at the *start* of the study, 1st September 2018.

As part of baseline summaries we will compare perinatal outcomes in the intervention and control arms during the time period before the intervention started in the intervention group, without providing P-values as any difference will be considered to be a chance finding.

For the primary statistical analysis, we will have the main strategy for analysis, as well as some sensitivity analyses based on intention-to-treat analysis populations, defined in the table below.

Table 2: Intention-to-treat populations

1	Baseline analysis of outcomes	Control (all baseline clusters, time starting 01.09.2018 to 31.09.2019)	Intervention (all baseline clusters, time starting 01.09.2018 to 31.09.2019)	Baseline analyses only including summary of prenatal
2	Primary statistical analysis	Control (all baseline clusters, time starting 01.09.2018)	Intervention (all baseline clusters, time starting 01.10.2019)	Baseline PLUS outcome analysis
3	Sensitivity statistical analysis 1	Control (all baseline clusters, time starting 01.10.2019)	Intervention (all baseline clusters, time starting 01.10.2019)	Baseline PLUS outcome analysis
4	Sensitivity statistical analysis 2	Control (all baseline clusters, time starting 01.10.2019)	Intervention (all baseline clusters, time starting 01.10.2019) + Pilot (all baseline clusters, time starting 12.12.2018)	Baseline PLUS outcome analysis
5	Sensitivity statistical analysis 3	Control (all baseline clusters, time starting 01.09.2018)	Intervention (all baseline clusters, time starting 01.10.2019) + Pilot (all baseline clusters, time starting 12.12.2018)	Baseline PLUS outcome analysis
6	Sensitivity statistical analysis 4	Control (all baseline clusters, time starting 01.10.2019) +1 year = starting 01.10.2020	Intervention (all baseline clusters, time starting 01.10.2019) + 1 year= starting 01.10.2020	Baseline PLUS outcome analysis

Per protocol populations

Changes in clusters and new clusters that arose during the study will be included in the per protocol analyses. These 'actual clusters' (as opposed to 'baseline clusters') are based on the cluster of residence of each case on the *day of birth*.

Clusters which incurred changes during the study are summarised in the 'Cluster Summary' table on the next page.

For the per protocol analyses, we will have a main strategy for analysis, as well as some sensitivity analyses based on per-protocol analysis populations, defined in the table below.

Table 3: Per protocol populations

Baseline analysis of outcomes	Control (actual clusters, time starting 01.09.2018 to 31.09.2019)	Intervention (actual clusters, time starting 01.09.2018 to 31.09.2019)
Main per protocol analysis	Control (actual clusters, time starting 01.09.2018)	Intervention (actual clusters, time starting 01.10.2019)
Sensitivity statistical analysis 1	Control (actual clusters, time starting 01.10.2019)	Intervention (actual clusters, time starting 01.10.2019)
Sensitivity statistical analysis 2	Control (actual clusters, time starting 01.10.2019)	Intervention (actual clusters, time starting 01.10.2019) + Pilot (actual clusters, time starting 12.12.2018)
Sensitivity statistical analysis 3	Control (actual clusters, time starting 01.09.2018)	Intervention (actual clusters, time starting 01.10.2019) + Pilot (actual clusters, time starting 12.12.2018)
Sensitivity statistical analysis 4	Control (actual clusters, time starting 01.09.2018)	Intervention (actual clusters, time starting 01.10.2019 + one year)

CLUSTER SUMMARY

Study	Cluster	Randomised	Randomisation procedure	Cluster changes	Change in study arm: direction	Change in study arm: timing	Effect on PLA intervention for CHWs	Effect on HC mentorship intervention	Cases affected	Analysis population
Pilot	A	No	Convenience sampling							ITT, PP
	N	No								ITT, PP
	O	No		Acquired 3 (/18) villages Nov'19 from cluster F	Yes: Control arm to Pilot arm	11 months after start of pilot	Yes: CHWs caught up with pilot PLA	No	281	ITT, PP
	Q	No		HP closed, no change to villages in cluster						ITT, PP
Intervention	B	Yes	Covariate constrained randomisation 03.02.19	8 (/24) villages removed: HP upgraded to HC and 8 villages moved to create new cluster V, June'19		Before intervention started				ITT, PP
	D	Yes								ITT, PP
	G	Yes								ITT, PP
	H	Yes								ITT, PP
	K	Yes		New HP opened, no change to villages in cluster						ITT, PP
	L	Yes								ITT, PP
	S	Yes		1 HP closed and 1 HP reassigned to new HC Jan'21: 15 (/77) villages removed to new cluster W		15 months after start of intervention				ITT, PP
	T	Yes		New HP opened, no change to villages in cluster						ITT, PP
	U	Yes	Coin toss June 2019 *	New cluster: new HC opened (Sept'18)						ITT, PP
	V	No	Kept in same study arm as origin cluster June 2019	*New* cluster: Cluster B HP upgraded to HC, acquired 8 villages from cluster B, June'19	No: Stayed in intervention arm	Before intervention started	No	No	360	PP
Control	C	Yes	Original covariate constrained randomisation 03.02.19							ITT, PP
	E	Yes								ITT, PP
	F	Yes		3 (/11) villages removed to cluster O, Nov'19						ITT, PP
	I	Yes								ITT, PP
	J	Yes								ITT, PP
	M	Yes								ITT, PP
	P	Yes								ITT, PP
	R	Yes								ITT, PP
	W	No	Defaulted to control arm January 2021	*New* cluster: new HC opened and acquired HP and 15 villages from cluster S, Jan'21	Yes: Intervention arm to Control arm	15 months after start of intervention	Yes: continued PLA in the 15 villages	Yes: stopped mentorship in HP	81	PP
		*cluster U existed at baseline but randomised later, separately from other baseline clusters. ITT intention-to-treat, PP per protocol								

DESCRIPTIVE STATISTICS

Broadly no change. All tables will exclude the pilot clusters.

BASELINE DATA

BASELINE CLUSTER CHARACTERISTICS

This table will now describe the baseline characteristics of the clusters in the intervention and control arms at the baseline and during the study. This will allow us to examine for cluster imbalance between intervention and control arms.

Table 4a: Cluster baseline covariates during each study time-period (as per figure 6).

** covariates included in randomisation procedure. # covariate included in choice of Pilot clusters*

Cluster	Population * n	Number of villages * n	Number of CHWs n	Number of primary healthcare staff n	Number of primary care facilities n	Distance from primary care facility to secondary care (referral) facility * (km)	Dry season travel time (mean travel time for each village to primary care facility) * (minutes)	Wet season travel time (mean travel time for each village to primary care facility) * (minutes)	Rurality (rural or urban)	etc
A Baseline Year 1 Year 2 Year 3 etc										

BASELINE CASES

No change except to note that the baseline control period (when both intervention and control arms are in the control condition) will be longer: 13 months rather than three months (figure 6). We may also look at the initial three-month 'true baseline control' period, which was when the pilot clusters were also in the control condition. This would remove all contamination risk from pilot to intervention clusters during the time when the intervention had started in pilot clusters, but not yet in the intervention arm.

Table 1b: Baseline cases (pregnancy outcome) and data completeness by cluster

Cluster	Deliveries n	Stillbirths n	Neonatal deaths n	Perinatal deaths n	Neonatal survival n	No information n
A	50	1	2	3	47	0
B	etc					
Control arm subtotal	etc					

OUTCOME DATA

CONSORT flow chart / Trial profile

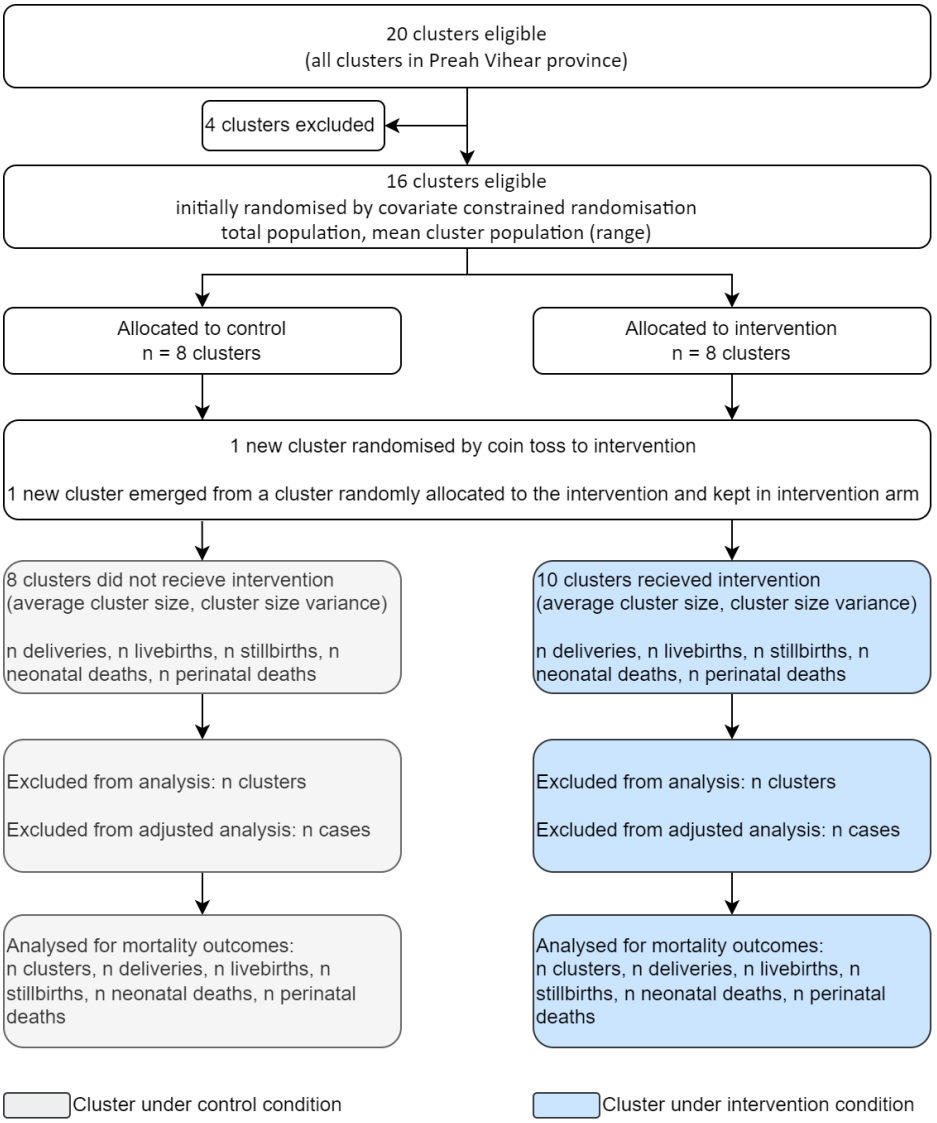

Figure 7: Cluster flowchart for the Saving Babies Lives cluster-randomised trial by allocated sequence and step. Cluster size = number of deliveries per month

Data completeness

No change except the data will not be presented by time-period (control, intervention).

Table 2: data completeness per cluster.

** Birth weight, gender, gestation*

Cluster	Deliveries n	Perinatal deaths n	Neonatal survival n	No information n (% of deliveries)	Perinatal deaths and complete covariates* n (% of perinatal deaths)	Perinatal survivals and complete covariates* n (% of perinatal survivals)
A	50	2	7	1 (%)	1 (50%)	30 (64%)
Control arm (subtotal) etc						

Delivery outcome data

Rather than presenting data by cluster and intervention condition (control, intervention), we will present the data by study arm (control and intervention arms) during years 1-3 of the study.

The cluster-level data will be presented in a similar format as a supplementary table.

Table 5: unadjusted and adjusted delivery outcomes (PMR, NMR, SBR) and patient-level characteristics by study arm

	Univariate			Multivariable		
Variable	Control	Intervention	P-value	Control	Intervention	P-value
Cases N deliveries **Outcome** Perinatal deaths n (PMR per 1000) Stillbirths n (SBR per 1000) Neonatal deaths n (NMR per 1000) Neonatal survival n Lost to follow up n **Gender** Male n (%) Missing n (%) **Gestation** Preterm n (%) Missing n (%) **Birth weight** n (median, IQR) Missing n (%) **Place of birth** Community n (%) Primary care facility n (%) Secondary care facility n (%) Other n (%) Missing n (%) **Place of death** Community n (%) Primary care facility n (%) Secondary care facility n (%) Other n (%) Missing n (%) etc						

Table 6: Births and deaths in intervention and control clusters at baseline and during the trial

	Baseline		Year 1		Year 2		Year 3		Year 2-3		Adjusted Year 2-3	
	Intervention	Control	Intervention	Control	Intervention	Control	Intervention	Control	Intervention	Control	Intervention	Control
Deliveries												
Live births												
Perinatal deaths												
Stillbirths												
Neonatal deaths												
Early neonatal deaths												
Late neonatal deaths												
Perinatal mortality rate												
Stillbirth rate												
Neonatal mortality rate												
Early neonatal mortality rate												
Late neonatal mortality rate												
etc												

Table 7: Comparison of mortality rates in intervention and control clusters

	Years 2-3		Adjusted Years 1-3		Adjusted Year 2-3		Adjusted Year 3	
	Odds ratio (95% CI)	P value	Odds ratio (95% CI)	P value	Odds ratio (95% CI)	P value	Odds ratio (95% CI)	P value
Perinatal mortality rate								
Stillbirth rate								
Neonatal mortality rate								
Early neonatal mortality rate								
Late neonatal mortality rate								
etc								

More detailed graphical representation of the secondary outcome (perinatal mortality) will be given. Below are some examples of how data might be presented. Survival between groups will also be described using Kaplan-Meier plots.

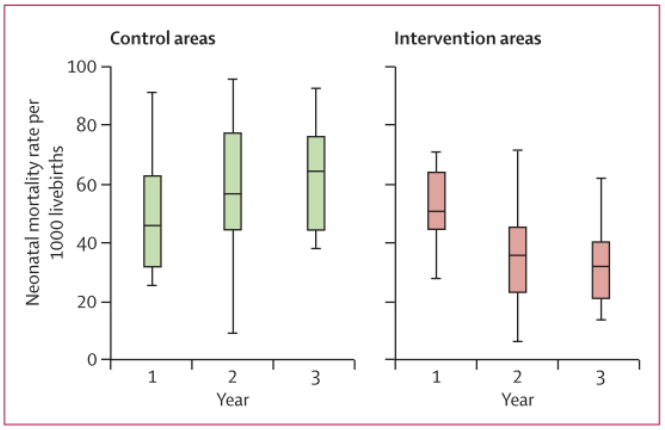

Figure 8: Example of boxplot of perinatal mortality rates by study arm (intervention, control) and study year (baseline, year 1, year 2, year 3) [copied from Azad et al 2010]

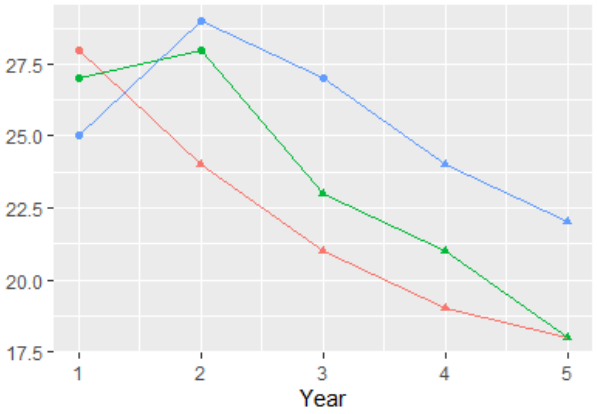

Figure 9: Trend in perinatal mortality rate (y axis) in pilot, control and intervention arms (colours) over time

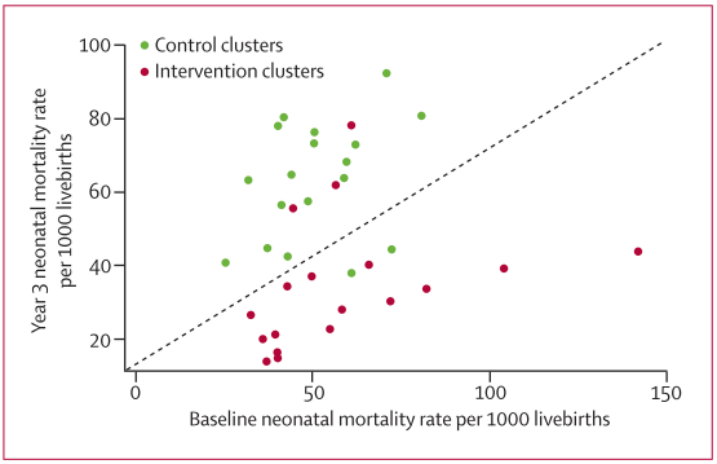

Figure 10: Example of scatterplot of perinatal mortality rates in year 3 (y axis) with rates at baseline (x axis) by study arm (intervention, control) [copied from Azad et al 2010]

Summary results of tertiary outcomes will also be presented in graphical or table format. For example:

Table 8: Quality of care indicators from the KAPES survey results in intervention and control clusters at baseline and during the study, presented by clinical goals. Cells relative to baseline score.

		Baseline		Year 1		Year 2		Year 3		Un / adjusted Year 2-3			
	Clinical goals	Intervention	Control	Intervention	Control	Intervention	Control	Intervention	Control	Intervention	Control	Odds ratio (95% CI)	P value
Knowledge	Neo resus												
	IPC												
	Routine care												
	Sick neo												
	Total												
Attitudes	Neo resus												
	IPC												
	Routine care												
	Sick neo												
	Total												
etc													

Table 9: Cause-specific mortality for all deaths during the study from verbal autopsy results

		Baseline		Year 1		Year 2		Year 3		Year 2-3		Un / adjusted Year 2-3			
		Intervention	Control	Intervention	Control	Intervention	Control	Intervention	Control	Intervention	Control	Intervention	Control	Odds ratio (95% CI)	P value
Stillbirth n (%)	In utero hypoxia														
	Congenital abnormality														
	Unknown														
Neonatal death n (%)	Birth asphyxia														
	Prematurity														
	Sepsis														
	Congenital abnormality														
	Other														

ANALYSIS

ANALYSIS OF PRIMARY OUTCOME

No change.

ANALYSIS OF SECONDARY OUTCOME

With a parallel CRT design, the modelling approach can be greatly simplified. A mixed effect Poisson regression model will be used to model PMR. A Negative Binomial Regression model will be considered if there will be overdispersion. The Incidence rate rations (IRR) along with the 95% confidence intervals will be reported. Both univariate and multivariable models will be fitted. Significance will be declared at 5% level of significance. We will include multi-level random effects to account for nested clustering within the data. The model will be adjusted for potential confounders, such as birth weight, gender, gestation.

Less consideration will be needed to check for time-point of actual crossover from control to intervention condition compared to the protocol because this is specific to SW-CRT designs.

ANALYSIS OF TERTIARY OUTCOMES

Count data will be compared using a multi-level Poisson model. Results will be presented as beta estimates with 95% CI. The levels to specify to look for variance (or clustering) may include village, facility, cluster.

Binary data will be compared using a multi-level mixed-effects logistic regression model. Results will be presented as relative risk with 95% CI.

Mortality outcomes

All tertiary mortality outcomes (facility PMR, overall NMR, VENMR) will undergo the same analysis as for the secondary analysis.

KAP/KAPES

The KAP and KAPES surveys measure quality of care at the primary facility-level. They involve several data collection methods (see separate excel document):

- K (KAP, KAPES): binary data (true/false), participant-level (CHW / healthcare worker)
- A (KAP, KAPES): 5-point likert scale (very unlikely/unlikely/neutral/likely/very likely), collected at participant-level (CHW / healthcare worker)
- P (KAP, KAPES): combination of response options (binary: yes/no, categorical, likert), collected at participant-level (CHW) or primary facility-level (by study team)
- E (KAPES): binary data, collected at primary facility-level (by study team)
- S (KAPES): combination of response options (binary: yes/no, categorical, count), collected at primary facility-level (by study team)

Analysis for quality of care outcomes (results from KAP and KAPES surveys) will be conducted at the primary facility-level per component (ie K/A/P/E/S) and overall (ie combined score). Summary results for each component will also be analysed by four clinically relevant topics: neonatal resuscitation, routine care, recognition of the sick neonate, infection prevention control. Survey data will be compared using ANCOVA or a mixed effects model (to handle longitudinal data), adjusted for baseline data.

Cost-effectiveness analysis

A detailed cost analysis will be carried out to estimate additional resources that are needed for the SBL programme over those required for standard care. These costs will be combined with estimates for the incremental cost of caring for neonates at facilities if attendance is found to increase, and by modelling subsequent survival benefits in terms of incremental disability-adjusted life years averted and quality-adjusted life years gained. Cost-effectiveness of the SBL programme compared to the normal standard of care (control arm) will be assessed.

REFERENCES

1. UN IGME. Levels & Trends in Child Mortality: Estimates: Report 2018. Who/Unicef/World Bank/Un [Internet]. 2018 [cited 2018 Nov 22];1–48. Available from: http://childmortality.org/files_v22/download/UN IGME Child Mortality Report 2018.pdf
2. Lawn JE, Blencowe H, Oza S, You D, Lee AC, Waiswa P, et al. Every Newborn: progress, priorities, and potential beyond survival. Lancet [Internet]. 2014 Jul 12 [cited 2018 Sep 4];384(9938):189–205. Available from: http://dx.doi.org/10.1016/S0140-6736(14)60496-7
3. Bhutta ZA, Das JK, Bahl R, Lawn JE, Salam RA, Paul VK, et al. Can available interventions end preventable deaths in mothers, newborn babies, and stillbirths, and at what cost? Lancet [Internet]. 2014 Jul [cited 2018 Sep 6];384(9940):347–70. Available from: http://dx.doi.org/10.1016/S0140-6736(14)60792-3
4. Darmstadt GL, Kinney M V, Chopra M, Cousens S, Kak L, Paul VK, et al. Who has been caring for the baby? Lancet [Internet]. 2014 Jul [cited 2018 Sep 6];384(9938):174–88. Available from: https://linkinghub.elsevier.com/retrieve/pii/S014067361460458X
5. Dickson KE, Simen-Kapeu A, Kinney M V, Huicho L, Vesel L, Lackritz E, et al. Every Newborn: health-systems bottlenecks and strategies to accelerate scale-up in countries. Lancet [Internet]. 2014 Aug [cited 2018 Sep 6];384(9941):438–54. Available from: https://linkinghub.elsevier.com/retrieve/pii/S0140673614605821
6. Bellad RM, Bang A, Carlo WA, McClure EM, Meleth S, Goco N, et al. A pre-post study of a multi-country scale up of resuscitation training of facility birth attendants: does Helping Babies Breathe training save lives? BMC Pregnancy Childbirth [Internet]. 2016 Dec 15;16(1):222. Available from: https://bmcpregnancychildbirth.biomedcentral.com/articles/10.1186/s12884-016-0997-6
7. Wrammert J, KC A, Ewald U, Målqvist M. Improved postnatal care is needed to maintain gains in neonatal survival after the implementation of the Helping Babies Breathe initiative. Acta Paediatr [Internet]. 2017 Aug 1 [cited 2020 Aug 6];106(8):1280–5. Available from: https://pubmed.ncbi.nlm.nih.gov/28316097/
8. National Institute of Statistics, Directorate General for Health, ICF International. Cambodia Demographic and Health Survey 2014 [Internet]. 2015 [cited 2019 Jan 16]. Available from: https://dhsprogram.com/pubs/pdf/fr312/fr312.pdf
9. Patel K, Say S, Leng D, Prak M, Lo K, Mukaka M, et al. Saving babies' lives (SBL) – a programme to reduce neonatal mortality in rural Cambodia: study protocol for a stepped-wedge cluster-randomised trial. BMC Pediatr [Internet]. 2021 Dec 7 [cited 2021 Sep 7];21(1):390. Available from: https://bmcpediatr.biomedcentral.com/articles/10.1186/s12887-021-02833-7
10. Hemming K, Taljaard M, McKenzie JE, Hooper R, Copas A, Thompson JA, et al. Reporting of stepped wedge cluster randomised trials: Extension of the CONSORT 2010 statement with explanation and elaboration. BMJ [Internet]. 2018 Nov 9 [cited 2019 Mar 28];363:k1614. Available from: http://www.bmj.com/lookup/doi/10.1136/bmj.k1614
11. World Health Organization. Verbal Autopsy Standards: The 2016 WHO verbal autopsy instrument. Manual for the training of interviewers on the use of the 2016 WHO VA instrument [Internet]. Geneva; 2017. Available from: https://www.who.int/standards/classifications/other-classifications/verbal-autopsy-standards-ascertaining-and-attributing-causes-of-death-tool
12. World Health Organisation. WHO | Verbal autopsy standards: ascertaining and attributing causes of death [Internet]. WHO. World Health Organization; 2016 [cited 2020 Sep 6]. Available from: http://www.who.int/healthinfo/statistics/verbalautopsystandards/en/
13. Allanson E, Tunçalp Ö, Gardosi J, Pattinson RC, Erwich JJHM, Flenady VJ, et al. Classifying the causes of perinatal death. Bull World Health Organ [Internet]. 2016 Feb 1 [cited 2020 Apr 27];94(2):79-79A.

Available from: https://doi.org/10.2471/BLT.15.168047

14. World Health Organisation. The WHO application of ICD-10 to deaths during the perinatal period: ICD-PM [Internet]. World Health Organization. 2016. Available from: http://www.who.int/reproductivehealth/projects/02-ICD-PM.pdf?ua=1

15. Waiswa P, Kallander K, Peterson S, Tomson G, Pariyo GW. Using the three delays model to understand why newborn babies die in eastern Uganda. Trop Med Int Heal [Internet]. 2010 Jul 14 [cited 2018 Sep 4];15(8):964–72. Available from: http://doi.wiley.com/10.1111/j.1365-3156.2010.02557.x

16. Kalter HD, Salgado R, Babille M, Koffi AK, Black RE. Social autopsy for maternal and child deaths: a comprehensive literature review to examine the concept and the development of the method. Popul Health Metr [Internet]. 2011 Dec 5 [cited 2020 Jan 29];9(1):45. Available from: http://www.pophealthmetrics.com/content/9/1/45

17. Moyer CA, Johnson C, Kaselitz E, Aborigo R. Using social autopsy to understand maternal, newborn, and child mortality in low-resource settings: a systematic review of the literature. Glob Health Action [Internet]. 2017 Jan 20 [cited 2020 Feb 9];10(1):1413917. Available from: https://www.tandfonline.com/doi/full/10.1080/16549716.2017.1413917

18. Thaddeus S, Maine D. Too far to walk: Maternal mortality in context. Soc Sci Med [Internet]. 1994 Apr;38(8):1091–110. Available from: https://linkinghub.elsevier.com/retrieve/pii/0277953694902267

19. World Health Organisation. Advocacy, communication and social mobilization for TB control: A Guide To Developing Knowledge, Attitude and Practice Surveys [Internet]. World Health Organisation. 2008 [cited 2020 Aug 30]. Available from: https://apps.who.int/iris/handle/10665/43790

20. Dron L, Taljaard M, Cheung YB, Grais R, Ford N, Thorlund K, et al. The role and challenges of cluster randomised trials for global health. Lancet Glob Heal [Internet]. 2021 May 1 [cited 2021 Oct 28];9(5):e701–10. Available from: http://www.thelancet.com/article/S2214109X20305416/fulltext

21. Hemming K, Haines TP, Chilton PJ, Girling AJ, Lilford RJ. The stepped wedge cluster randomised trial: rationale, design, analysis, and reporting. BMJ [Internet]. 2015 Feb 6 [cited 2019 Mar 28];350(feb06 1):h391–h391. Available from: http://www.bmj.com/

22. Manandhar DS, Osrin D, Shrestha BP, Mesko N, Morrison J, Tumbahangphe KM, et al. Effect of a participatory intervention with women's groups on birth outcomes in Nepal: cluster-randomised controlled trial. Lancet [Internet]. 2004 Sep 11 [cited 2020 Aug 6];364(9438):970–9. Available from: https://pubmed.ncbi.nlm.nih.gov/15364188/

23. Woertman W, De Hoop E, Moerbeek M, Zuidema SU, Gerritsen DL, Teerenstra S. Stepped wedge designs could reduce the required sample size in cluster randomized trials. J Clin Epidemiol [Internet]. 2013 [cited 2020 Oct 23];66:752–8. Available from: http://dx.doi.org/10.1016/j.jclinepi.2013.01.009

24. R: The R Project for Statistical Computing [Internet]. [cited 2020 Aug 30]. Available from: https://www.r-project.org/

25. Barker D, McElduff P, D'Este C, Campbell MJ. Stepped wedge cluster randomised trials: a review of the statistical methodology used and available. BMC Med Res Methodol [Internet]. 2016 Dec 6 [cited 2020 Jul 27];16(1):69. Available from: http://bmcmedresmethodol.biomedcentral.com/articles/10.1186/s12874-016-0176-5

26. Hussey MA, Hughes JP. Design and analysis of stepped wedge cluster randomized trials. Contemp Clin Trials. 2007 Feb;28(2):182–91.

27. Hemming K, Girling A. A menu-driven facility for power and detectable-difference calculations in stepped-wedge cluster-randomized trials. Stata J. 2014 Jun 1;14(2):363–80.

28. Campbell MK. Consort 2010 statement: extension to cluster randomised trials. 2012 [cited 2022 Mar 1]; Available from:

http://www.bmj.com/permissionsSubscribe:http://www.bmj.com/subscribeBMJ2012;345:e5661doi:10.1136/bmj.e5661