Statistical analysis plan for Safer Births Bundle of Care, phase II

Version: 1.0 Date: October 28, 2025

This Statistical Analysis Plan (SAP) describes analytical approaches for the assessment of intervention effects on perinatal and maternal (clinical) outcomes of the Safer Births Bundle of Care (SBBC) project, phase II (registered as ISRCTN92381311), and for exploring promotors and inhibitors for change. It was finalized before completion of the data, but after some preliminary, descriptive analyses on partial data, including observed rates of outcome over time and assessment of missingness.

1 Introduction

SBBC phase II is a quality improvement study, implemented in 142 facilities in five regions in Tanzania. These facilities comprise all CEmONC (Comprehensive Emergency Obstetrics and Newborn Care) facilities in these five regions. Among the facilities, 30 took part in SBBC phase I and 112 facilities are new. As such, SBBC phase II represents an upscaling but also introduced new elements vis a vis SBBC phase I, see the protocol at https://www.isrctn.com/ISRCTN92381311 for details.

Identification of eligible facilities and central preparations started in December 2023. From January 2024 preparations for the implementation took place at the individual facilities, and the intervention started April-June 2024.

2 Data

Baseline data from 2023 and the first months of 2024 were collected retrospectively from prospectively registered births in hospital books for the 112 new facilities. For the 30 facilities continuing from SBBC phase I, baseline data from this period were gathered prospectively via the data system set up for SBBC phase I. From March 2024, designated data collectors visited all facilities regularly i.e. at least biweekly, registering data on newly recorded births from hospital books.

Recorded information from hospital books included outcomes and clinical variables (mother's age and parity, child's gender, birth weight, number of fetuses, etc.) related to the child and the mother. Every child and every birth have a unique ID number. Unfortunately, repeated births by the same mother cannot be identified.

Prospective collection of additional clinical data (source of admission, fetal heart rate, device use, etc.) was done separately starting June 2024 (fully operable from August 2024), i.e., only in the intervention period, and not linked to the individual births. Data about training practices at each facility were recorded by a separate digital system (LIFT) related to the training devices. Other facility related information (staff density) were collected from hospital management. Readiness data at baseline were collected using a standardized WHO form (Juma et al, 2024).

Data collection on all facilities will be running until end 2025. More details about the data gathering systems are given in the protocol and the data management plan at https://www.isrctn.com/ISRCTN92381311. Descriptions of outcome variables and covariates relevant for the analyses described in the analysis plan are given below.

2.1 Outcomes

Two primary outcomes were listed in the protocol, and are defined as:

- Perinatal death defined as intrapartum stillbirth (ISB) or newborn death within the first 24 hours of life.
- Maternal death within seven days postpartum.

Unfortunately, death within 24 hours is not directly available in the collected data from hospital books, apart from for newborns admitted to NICU. For all others, we have status at discharge and date (but not time of day) of discharge. We will define "early newborn death" as discharged as dead within (i.e., including) the date of birth plus one day. Which means that, depending on the time of day of the birth, death may have occurred at most <48 hours after birth.

Similarly, seven days maternal mortality will be defined as being dead at discharge if date of discharge for the mother is at most seven days after the date of delivery. (If date of discharge is missing for mother, we will impute it using date of discharge for child.)

For the baseline data from SBBC-I facilities, we do have the 24-hour newborn deaths and the seven-days maternal deaths directly available. However, to make the rates comparable, we will apply the same definitions to these data as well.

For perinatal death, newborns that are referred out during the first 24 hours will be regarded as having unknown outcome. Antepartum/macerated stillbirths (ASB) will be excluded from the sample. For maternal death, mothers referred out before seven days will be considered to have unknown outcome. ASB will be included and used as an adjustment variable when considering maternal outcome.

The following secondary outcomes were defined:

- 24-hour newborn death.
- Intrapartum stillbirth (ISB).
- 7-days perinatal death (including ISB).

The 24-hour newborn deaths will be redefined as "early newborn deaths" as above. Seven-days newborn deaths will be assessed using status at discharge and date of discharge as for maternal outcome.

2.2 Covariates

In Table 1 an overview of covariates considered relevant for the clinical outcomes are listed. These variables are considered to be the most relevant prognostic factors, without too serious reliability and missingness issues.

Table 1: Overview of covariates relevant to include in the modelling

Covariate	Available period (baseline/intervention)	Available at level (individual/facility)	Fixed or time-varying	Details	Comment
Child					
Birth weight	Both	Individual	Time-varying	Grams	
Sex	Both	Individual	Time-varying	Male, female, ambiguous, or unknown	Ambiguous and unknown regarded as missing
Mother/pregnancy					
Age	Both	Individual	Time-varying	Years	
Parity	Both	Individual	Time-varying	Mother's number of previous births	
Multiples	Both	Individual	Time-varying	Number of fetuses	Dichotomize into one or more than one
ASB	Both	Individual	Time-varying		Adjustment variable for maternal outcome; ASB excluded for perinatal/newborn outcomes
Source of admission[1]	Intervention	Facility	Time-varying	Home, referral, or unknown → monthly proportions of referral-ins (unknown disregarded)	
Fetal heart rate (FHR) at admission[1]	Intervention	Facility	Time-varying	Heard/present, not heard/not present, or not measured/unknown → monthly proportions of not heard/present (not measured/unknown disregarded)	
Term/preterm[1]	Intervention	Facility	Time-varying	Term or preterm → monthly proportions of preterm	
Use of Moyo[1]	Intervention	Facility	Time-varying	Yes, no, or unknown → monthly proportion of births where Moyo used	
Facility					
Region	Both	Facility	Fixed	Five regions	

Level of health facility	Both	Facility	Fixed	Health center, district hospital, or regional referral hospital	
Staff density	Both	Facility	Fixed	Number of healthcare workers at the facility at the start of the intervention period, divided by the average number of births per month at the facility during the intervention period, multiplied by 100	Assume representative for the whole study period
Readiness level	Baseline	Facility	Fixed	0-100	
Training data – Team training	Intervention	Facility	Time-varying	Monthly amount of team scenario training (average 5 persons involved) divided by number of staff. Assume effect lagged by one month. Also consider cumulative amount of training.	
Training data – Individual skill training	Intervention	Facility	Time-varying	Monthly amount of individual skill training divided by number of staff. Assume lagged effect (one month). Also consider cumulative amount of training.	
Mentorship visits	Intervention	Facility	Time-varying	Number of mentorship visits. Assume lagged effect (one month). Also consider cumulative number of visits.	Number of mentors larger for larger units for the planned visits, but not for the unplanned ones.

[1)] These variables were collected on a separate form for each birth during most of the intervention period (starting June 2024) but not linked to the other information about the individual births, thus can only be used as facility level variables.

3 (Main) Analysis

3.1 Descriptive

Basic tabular summaries of outcomes, clinical variables, and process variables will be presented, stratified by the two intervention periods (baseline, intervention). These tables will be made separately for facilities that was part of SBCC phase I and the new hospitals. Further, tables stratified by region and hospital level will be made.

3.2 Estimation of overall efficacy

Logistic regression will be used to assess the overall efficacy of the intervention, including data from all 142 facilities. We anticipate varying intervention effects per region and per facility level and depending on whether the facility was part of SBBC phase I or not, and possibly also varying intervention effects between combinations of these factors. Further, we anticipate time-varying intervention effects as well as clustering on two levels, birth (for multiples) and facility.

Thus, the full model will be a mixed-effects logistic regression model, incorporating the following:

Fixed effects:

a) Time-varying intervention effects modelled using a continuous variable representing time since intervention start (set to zero for all baseline observations). To capture potential non-linear effects, time will be modelled using splines. The spline-transformed time variable, $f(t)$, will be interacted with an indicator for intervention status (intervention vs. baseline), i.e., $I(\text{intervention})$.
b) Region-specific treatment effects, modeled through an interaction between region (five regions) and the intervention-by-time spline term, $I(\text{intervention}) \cdot f(t)$.
c) Facility level-specific treatment effects, modeled through an interaction between facility level (three levels) and $I(\text{intervention}) \cdot f(t)$.
d) Differing treatment effects between facilities that were part of SBBC phase I and those new in SBBC phase II, through interaction between SBBC-I (yes/no) and $I(\text{intervention}) \cdot f(t)$.
e) All higher level interactions between region, facility level, SBBC-I, and $I(\text{intervention}) \cdot f(t)$.

Random effects:

f) Random effects (random intercepts and random intervention effects) for facilities and for births nested within facilities.

In Stata syntax, the full regression model looks like this:

```
melogit outcome ///
  i.region i.facility_level i.sbbc_phase1 i.intervention ///
  i.intervention#c.s_time ///
  i.region#i.intervention#c.s_time ///
  i.facility_level#i.intervention#c.s_time ///
  i.sbbc_phase1#i.intervention#c.s_time ///
  i.region#i.facility_level#i.intervention#c.s_time ///
```

i.region#i.sbbc_phase1#i.intervention#c.s_time ///

i.facility_level#i.sbbc_phase1#i.intervention#c.s_time ///

i.region#i.facility_level#i.sbbc_phase1#i.intervention#c.s_time ///

|| facility_id: i.intervention i.intervention#c.s_time ///

|| birth_id: i.intervention i.intervention#c.s_time ///

, covariance(unstructured)

In the (likely) case of convergence problems, the model will be reduced in the following order:

1. Reducing complexity in the random effects.
2. Reducing complexity in the fixed effects by removing higher order interactions.
3. Reducing complexity in the fixed effects by reducing the degrees of freedom for the splines of time, ultimately assuming linear effect of time.
4. After each reduction in fixed effects, try in the random effects again.

Regarding 1: Experience from SBBC phase I was that the clustering on birth level (for multiples) was difficult to account for without convergence issues. If the same problem is encountered here and not resolved by removing random intervention effects at two levels, we will resort to the same solution as used in phase I, i.e., including only one of the babies in births with multiples.

Once a model has been identified that converges successfully and shows no signs of numerical instability, we will proceed to reducing the fixed effects structure by optimizing Akaike's Information Criterion (AIC). Specifically, any parameter whose removal results in a reduction of the AIC by at least 2 will be excluded from the model. If possible, random effects will be reintroduced.

In SBBC-I, FHR at admission was by far the most important individual-level adjustment variable for perinatal deaths, and for maternal deaths, FHR at admission and ASB were most important. Referral-in (yes/no) was also of some importance. FHR at admission and referral-in are only available in the intervention period here and only on facility level (FHR also in a less informative variable than in SBBC-I), and this is also the case with preterm (yes/no). Thus, in addition to ASB (for maternal outcome), we will in the current analyses have to suffice with adjusting for birth weight (categorical including "missing" as a category), mother's age (non-linear effect), and multiples (yes/no), which all were found to have some adjustment effect in SBBC-I. Since we do not expect changes in the population during the study period, these adjustments will be mostly relevant for comparisons between regions, facility levels, and SBBC-I (yes/no), where changes in referral patterns among the participating facilities during the intervention may play a role.

The overall effect of intervention will be presented as the marginal relative risk (RR) at 1.5 years after start intervention, obtained using g-computation (Snowden et al, 2011). Maximum likelihood estimation will be used for model fitting, the final model will be fitted using restricted maximum likelihood.

3.3 Statistical process control to study time to effect

Cumulative sum chart (CUSUM) procedures will be used to monitor for changes in each of the outcome variables throughout the intervention period. The CUSUMs will be based on logistic regression models estimated from the baseline data (Steiner et al, 2000). An overall CUSUM plot, risk-adjusted for region, facility level, and whether a facility was part of SBBC phase I or not will be

constructed. In addition, separate plots will be constructed per region, facility level, and SBBC-I (yes/no), as we anticipate substantial differences both in baseline levels and intervention effects between the levels of these factors. The signal limit for the CUSUM will be set according to a probability of 5% of getting a false alarm during 100 000 births in case there is no change from the baseline.

3.4 Exploration of contextual and implementation fidelity predictors of outcome trends during the intervention period

To explore facility-level factors predicting change in clinical outcomes during the intervention period, we will analyze monthly aggregated data from each facility using mixed effects Poisson regression. Random effects will be used to model dependency in data from the same facility. Development of outcome rates during the intervention period will be modelled non-linearly using splines and accounting for the number of births in each facility in each month. Candidate time-fixed predictor variables are region, facility level, average no. of births per month in the baseline period, baseline levels of outcomes, staff density, and readiness level. Effects of these will be assessed by interaction terms between the variable and the splines functions of intervention month. Effects of time-varying predictors like degree of uptake of new routines (use of Moyo and NeoBeat), and (cumulative) amount of training and mentorship involvement will be assessed by their rate ratios, assuming lagged (by one month) effects of training and mentorship involvement.

Adjustments will be made for monthly rates of referrals-in, births with not heard/detected FHR, and preterm births for each facility. Since these, as well as data regarding use of Moyo and NeoBeat, are only available for a substantial proportion of births from August 2024 and onwards, adjusted results and results for the latter predictor variables will be presented for the period August 2024 to December 2025. Unadjusted results, and results adjusted for BW, mother's age, and multiples, will be presented for the whole intervention period whenever possible.

In these analyses, the 30 SBBC-I facilities and the 122 new facilities will be analyzed separately.

4 Reporting and multiplicity considerations

Due to the non-randomized pre-post design of the study no causal claims will be made based on the estimated effects/associations.

All estimates will be presented with 95% confidence intervals (CI) and corresponding p-values from significance tests. For the estimates of overall efficacy for the two co-primary outcomes, p-values <0.025 will be considered statistically significant. All other p-values will be interpreted as exploratory.

If a confidence interval goes beyond the region of possible values for the parameter, we will report the interval truncated to possible values.

5 Graphical displays of results

Results from the main analyses will be supplemented with prediction plots based on the final models.

Furthermore, to accompany the CUSUM plots and illustrate the impact of the intervention over time, variable life adjusted displays (VLAD) will be made (Lovegrove, 1997). These plots will be made by aggregating the difference between the estimated conditional probability for fatal outcome based on a model estimated from baseline data, and the observed outcome, and can thus be interpreted in terms of estimated numbers of lives saved. The same logistic regression models as used for the CUSUM plots described in Section 3.3 will be used to estimate the probability of fatal outcome under baseline conditions.

For the promotor/inhibitor analyses, estimated marginal rates will be presented over time for each considered factor, based on the Poisson regression models.

6 Missing data and sensitivity analyses

The main efficacy analysis will be based on the available cases, assuming missing completely at random (MCAR) missingness in covariates (preliminary assessment of amount of missingness <1%) and missing at random (MAR) missingness in the outcome variables. We do expect some missingness in the outcome variables. For assessing sensitivity to different assumptions regarding missing data in the primary outcomes, perinatal death and 7-days maternal death, we provide sensitivity analyses based on missing not at random (MNAR) best/worst-case scenarios:

- As best-case scenarios we assume all missing being non-fatal outcomes.
- As worst-case scenario for perinatal death, we assume the following missing being fatal outcomes: referred-out cases; birth by CS; BW<1500 grams or missing; those requiring resuscitation using ventilation; low Apgar score (i.e., <7) at 5 minutes, or, if missing at 5 minutes, Apgar score <7 at 1 minute. The rest are assumed non-fatal cases.
- For maternal outcome we assume all missing with antepartum stillbirths, eclampsia, labour complications (3-degree tear, retained placenta, postpartum bleeding, blood transfusion), and referred-out cases to be fatal outcomes in the worst-case scenario.

7 Supplementary and supporting analyses

As a supplement to the main results in 3.2, marginal RRs will be presented according to combinations of factors (region, facility level, SBBC-I) and for different time points in the intervention period.

Although the intervention does not include means to reduce ASB, and does not target directly 2-7 days mortality, these outcomes are adjacent to the primary perinatal outcome and may be influenced indirectly, either due to changes in reporting or through possible postponement of the death event. Thus, it is of interest to assess potential changes in these indicators as well. For this, we will provide plots showing trends in rates of ASB and 2-7 days newborn deaths over the study period.

For the new SBBC-II sites, we have two sources of baseline data, each with its own limitations and merits. The retrospective data from 2023 and two first months of 2024 are not affected by the intervention in any way and represent a large sample. On the other hand, the data quality is expected to be lower than for the prospective data and with under-reporting of mortality. The prospective baseline data were collected in a period with preparation for the intervention, training of champions, probably increased awareness of optimal birth care, and possibly a Hawthorne effect related to the awareness of being observed/"measured". The data quality is more similar to that of the data collected during the intervention period, but the sample size is smaller than for the retrospective baseline data.

The relative effects of the expected more complete reporting of poor outcomes and the impact of the preparations in the prospective baseline data are unknown, but we believe that the underreporting in the retrospective data is substantial. We will present some analyses comparing the prospective and the retrospective baseline data, to inform the discussion of results of the main analyses.

8 Software

The analyses will be performed in Stata version 18 and R version 4.5.1, or newer versions.

9 References

ISRCTN92381311. Scaling up safer births bundle of care - keeping more mothers and babies safe during and after birth. ISRCTN Registry; 2025. Available at: https://www.isrctn.com/ISRCTN92381311.

Juma D, Stordal K, Kamala B, *et al*. Readiness to provide comprehensive emergency obstetric and neonatal care: a cross-sectional study in 30 health facilities in Tanzania. *BMC Health Serv Res* 2024;**24**:870. DOI: 10.1186/s12913-024-11317-0.

Kamala BA, Ersdal HL, Moshiro RD, *et al*. Outcomes of a Program to Reduce Birth-Related Mortality in Tanzania. *N Engl J Med* 2025;**392**(11):1100-1110. DOI: 10.1056/NEJMoa2406295.

Kamala BA, Ersdal HL, Mduma E, *et al.* SaferBirths bundle of care protocol: a stepped-wedge cluster implementation project in 30 public health-facilities in five regions, Tanzania. *BMC Health Serv Res* 2021;**21**:1117. DOI: 10.1186/s12913-021-07145-1.

Kamala BA, Ersdal H, Moshiro R, *et al*. Improvements in obstetric and newborn health information documentation following the implementation of the Safer Births Bundle of Care at 30 facilities in Tanzania. *Healthcare* 2024;**12**(3):319. DOI: 10.3390/healthcare12030319.

Lovegrove J, Valencia O, Treasure T, *et al*. Monitoring the results of cardiac surgery by variable life-adjusted display. *Lancet* 1997;**350**(9085):1128-1130. DOI: 10.1016/S0140-6736(97)06507-0.

Snowden JM, Rose S, Mortimer KM. Implementation of g-computation on a simulated data set: demonstration of a causal inference technique. *Am J Epidemiol* 2011;**173**(7):731-738. DOI: 10.1093/aje/kwq472.

Steiner SH, Cook RJ, Farewell VT, Treasure T. Monitoring surgical performance using risk-adjusted cumulative sum charts. *Biostatistics* 2000;**1**(4):441-52. DOI: 10.1093/biostatistics/1.4.441.