Design considerations for a cluster randomized controlled trial to evaluate the impact of an Optimized Intervention Package on maternal and child health outcomes in Rwanda

27 August 2024

Contents

1. Background

1.1. Primary goal and research questions of TIP

The Government of Rwanda (GOR) is interested in improving maternal and child health, with a focus on stunting. Stunting is common in Rwanda, with 33% of under-5-year-old children considered stunted. Given the acuteness of the challenge, there is interest from GOR and international stakeholders in partnering with TIP Global Health to use E-Heza—a digital health platform to augment health system delivery—as a mechanism to facilitate delivery of high quality maternal and child health services in specified districts of Rwanda where stunting rates are high.

This offers an opportunity to design a rigorous study to quantify the impact of an optimized intervention package on health and health-system outcomes. Specifically, the evaluation study will test the hypothesis that the E-Heza facilitated intervention package improves maternal and child health outcomes in pregnancy, at birth, and during the first two years of the child's life. Secondarily, the evaluation will also quantify and build an understanding of the effect of this package on the performance of the health system.

1.2. Background

Linear growth faltering, also known as stunting, is common in Rwanda. On a national level, a third of under-5-year-old children are considered stunted (i.e., have length-for-age that is more than 2 standard deviations below the reference mean). An even higher prevalence (47%) is found in the poorest tertile (DHS 2019-2020). Stunting is associated with increased lifelong risk for morbidity and mortality and many other problems that contribute to loss of human and economic capital in the society. Although prevention is difficult, recent evidence suggests that it is possible if the intervention starts early—during the child's fetal period, is multisectoral, and continues for the first two years of the child's life. This is where improved quality of maternal and child health can play a distinct role.

Since its founding in 2008, TIP Global Health has been tackling challenges faced by local primary health care facilities in Rwanda. By partnering with communities, TIP Global Health develops effective and self-sustaining approaches to reduce the burden on health workers and strengthen the quality of care they provide. One promising approach TIP Global Health has developed in this regard is the E-Heza digital health platform. E-Heza was originally developed to strengthen the effectiveness of Childhood Growth Monitoring Programs in rural Rwanda by facilitating group health assessments to promote screening, early diagnosis, treatment, and referral for cases of childhood undernutrition, and to make it possible to provide personalized health education to mothers based on their child's growth trends. This model was later expanded to serve the entire spectrum of maternal and child health, utilizing data trends to provide personalized care, improve maternal engagement in care, celebrate successes, and identify challenges before they become emergencies, while preventing frontline health worker burnout by decreasing workloads, supporting workflows, and maximizing opportunities for meaningful interactions with patients. TIP Global Health hypothesizes that higher quality care not only directly improves health outcomes but also leads to increased maternal engagement in care, which is a second pathway through which the approach contributes to improved health outcomes.

Informed by a deep understanding of Rwanda's local health systems and workflows, E-Heza was developed to optimize implementation of the Rwanda Ministry of Health's best practice clinical protocols. The Government of Rwanda and research partners would like to use E-Heza as part of a rigorous evaluation of the impact of their optimized package of interventions on childhood stunting.

1.3. Purpose of the memo

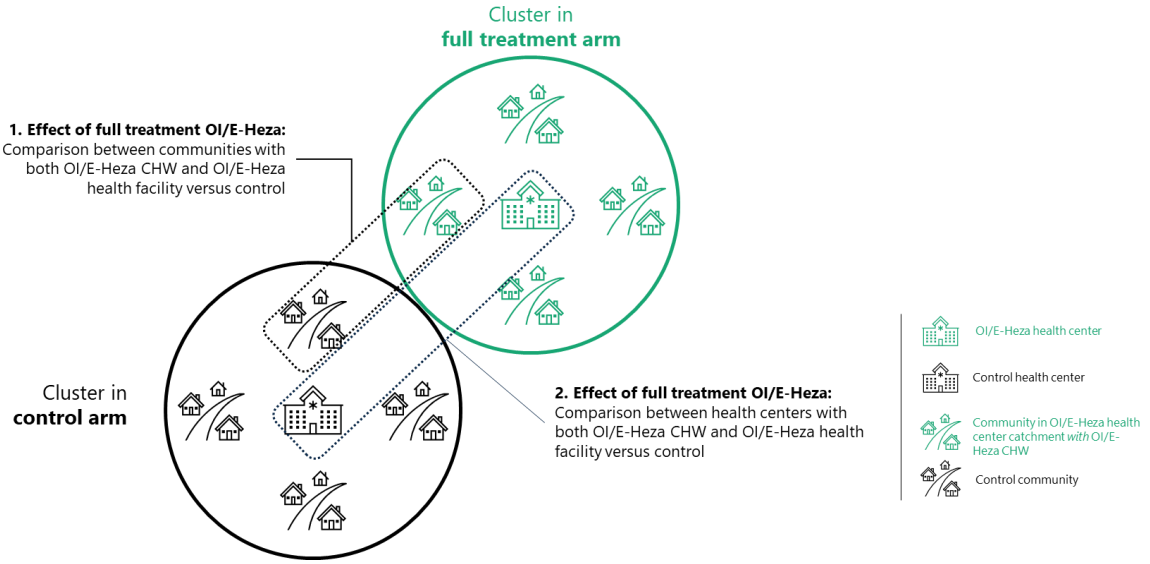
This memo describes the design of a cluster randomized controlled trial (RCT) to evaluate and understand the impacts of an optimized intervention (OI) package on maternal and child health delivered through E-Heza. We refer to this as the OI/E-Heza package throughout this memo. We discuss the benefits and limitations of the design and the associated data collection needed, including an outline of the population of interest, key outcomes to measure, timeframes and sample size requirements.

2. Evaluation design

Below we outline a two-armed RCT design, which has a control arm and a treatment arm (**Figure 1**). The RCT will be implemented across available health facilities and their catchment populations; in the language of the evaluation, the health facilities are our clusters. We outline different scenarios for the number of clusters and sample size implications for each below. In each cluster, comparing community members and health centers in the control arm with those in the treatment arm will give us an estimate of the effect that the combination of both health facility staff and community health workers (CHWs) receiving OI/E-Heza training and equipment has on outcomes.

While our primary interest is in health impacts on the community, we will glean considerable insights by comparing health facilities too. Coupled with a cost-effectiveness analysis, the results from the RCT will provide evidence on the OI/E-Heza package's potential as an approach to reduce stunting, improve health outcomes in general, and improve the performance of the health system.

Figure 1. *Design for a randomized controlled trial of the effects of the optimized intervention delivered through E-Heza at facilities and by community health workers*



Importantly, implementers will need to ensure that those health centers and their catchment of communities allocated to the treatment are fully enabled with the OI/E-Heza. It will also be important that control sites do not receive the OI/E-Heza package during the study period. Any other initiatives or interventions that may occur (outside the scope of this study) should be the same in both treatment and control sites, and it will be important that these are well documented.

3. Data collection and sample size

This evaluation will seek to understand the impacts of the OI/E-Heza package across three types of outcomes:

1. Maternal, infant, and child health outcomes among mother-child dyads in target age groups
2. Health system performance outcomes
3. Cost-effectiveness modelled as cost of implementation per unit outcome

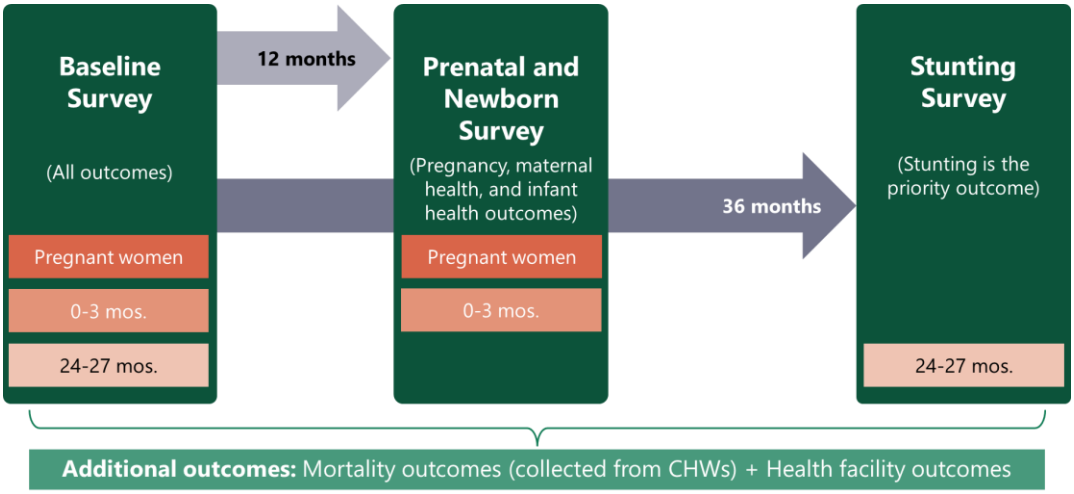
To ensure consistency and comparability of data across treatment and control sites, the data collection plan will use third-party anthropometric measurement and surveys. Additionally, it will seek to use E-Heza and health facility data where available.

3.1. Study approach and participants

We propose a repeated cross-sectional study to test the hypothesis that the OI/E-Heza package improves maternal and child health outcomes in pregnancy, delivery and during the first two years of the child's life. Mother-baby dyads in three age groups will be sampled: women during pregnancy, children aged 0-3 months, and children 24-27 aged months.

As depicted in **Figure 2**, the design entails three surveys: a baseline survey for all outcomes, a survey conducted 12 months after baseline to assess maternal and infant health outcomes, and a final survey 36 months after baseline to assess stunting. Additionally, maternal, infant, and neonatal mortality will be tracked throughout the study, as we discuss in further detail ahead. For each survey, study participants will be identified through a community-level census of all households with a pregnant woman and/or children in the target age ranges in each health facility catchment area i.e., in each study cluster.

Figure 2. *Cross-sectional study design components*



Cross-sectional approaches offer several advantages over longitudinal studies in the context of child health and nutrition evaluations (Menon, Rawat, and Ruel 2013). First, they can have flexible timelines. This allows them to either extend the evaluation endline or otherwise align data collection intervals more easily with implementation timelines, which may shift. This is important in the context of child nutrition evaluations to avoid the risk of children aging out of the age range at which outcomes are relevant if there are implementation delays, or conversely of being measured too early without having had full exposure to the intervention.

Second, cross-sectional approaches also avoid the issue of sample attrition, which is an important challenge in longitudinal surveys. Several longitudinal surveys in sub-Saharan Africa have lost between 3 and 23 percent of their sample per year (Alderman et al. 2001). Mathematica also recently faced this challenge in Rwanda, where significant time and resources were required to trace people who had been enrolled in an implementing organization's earlier study, and many of whom had migrated and were unreachable. Sample attrition is often non-random, which can introduce bias into findings, while efforts to trace participants can increase study costs. If too much of the sample is lost to follow-up, this can reduce the statistical power to detect significant effects and limit generalizability of results.

In summary, we recommend a cross sectional approach because 1) it will ensure that any changes to implementation timelines will not compromise the evaluation, and 2) it avoids the likely loss to follow up that can bias or impede detection of significant results.

3.2. Study outcomes

Key outcomes of interest, definitions, and data sources for each are listed in **Table 1**. Outcomes include maternal and child health impacts that will be measured among pregnant women and mother-child dyads in given age ranges. They also include outcomes to gauge aspects of health system performance.

Table 1 also presents the age groups for which outcomes will be measured. As mentioned above, those age groups are pregnancy, children aged 0-3 months, and children 24-27 aged months. For logistical feasibility and because the evaluation will rely on direct measurement for most outcomes, age groups are ranges, rather than single specific ages. These age ranges are designed to be narrow enough to allow the evaluation to compare children at similar developmental stages and to align with the established age ranges for various MNCH interventions, but wide enough that it is possible to identify a sufficient number of children in each cluster. Age ranges narrower than those proposed may require additional efforts to identify children within each age range, which can be resource-intensive and time-consuming, especially if the population of interest is small.

***Table 1**. Outcomes and measurement*

Primary Outcomes[1]	Definition	Source	Age group	Survey		
				Baseline	Prenatal & newborn	Stunting
Hemoglobin (Hb) in pregnancy	Pregnant women with < 11 g/L (anemia) Pregnant women with Hb> X g/L (high Hb)[2]	Direct measurement: blood tests, health facility/ CHW records	Pregnancy	**X**	**X**	*optional*
1st trimester ANC	Women attending ANC 1 at < 14 weeks' gestation	Survey/ confirmation via health records	Pregnancy	**X**	**X**	*optional*
Low birthweight	Children with weight at birth <2,500 g	Survey / confirmation via health records	0-3 mos.	**X**	**X**	*optional*
4+ ANC visits	Women attending at least 4 ANC visits	Survey/ confirmation via health records	0-3 mos.	**X**	**X**	*optional*
Stunting / Severe stunting	Children with height for age (HAZ) <-2 SD Children with height for age (HAZ) <-3 SD	Direct anthropometric measurement	24-27 mos.	**X**	NA[3]	**X**
Neonatal mortality	Deaths during the period birth to 28 days	Health facility/CHW records	*Within given time frame (see **Annex 3**)*	*Ongoing CHW record keeping*		
Infant mortality	Deaths between birth to < 12 mos.	Health facility/CHW records	*See **Annex 3***	*Ongoing CHW record keeping*		
Secondary Outcomes						
(Pre)eclampsia	Women diagnosed with (pre)eclampsia Women diagnosed with (pre)eclampsia who receive proper treatment[4]	Survey/ confirmation via health records	Pregnancy and 0-3 mos.	**X**	**X**	*optional*
Post-partum hemorrhage	Women with severe blood loss[5] within 24 hours after birth	Survey/ confirmation via health records	0-3 mos.	**X**	**X**	*optional*
Preterm birth	Births at < 37 gestational weeks	Survey/ confirmation via health records	0-3 mos.	**X**	**X**	*optional*
Small for gestational age	Children with birthweight < 10th percentile for gestational age	Survey/ confirmation via health records	0-3 mos.	**X**	**X**	*optional*
PNC attendance	Women attending PNC within 6 weeks	Survey/ confirmation via health records	0-3 mos.	**X**	**X**	*optional*
Medication / Supplementation	Women / children receiving food supplements Women / children receiving micronutrient supplements (iron/folic acid, calcium)	Survey / confirmation via health records	Pregnancy 0-3 mos. 24-27 mos. (if relevant)	**X**	**X**	*If relevant*

	Women / children receiving deworming treatment					
Vaccinations	Children receiving all age-appropriate vaccinations[6]	Survey / confirmation via health records	24-27 mos.	**X**	**X**	**X**
Maternal mortality	Deaths in pregnancy to 42 days postpartum	Health facility/CHW records	*See **Annex 3***	*Ongoing CHW record keeping*		
Still birth	Pregnancies ending in fetal death at or after 28 weeks of gestation	Health facility/CHW records	*See **Annex 3***	*Ongoing CHW record keeping*		
Malaria prevention	Dyads using insecticide treated bed nets (ITNs)	Survey	All age groups	**X**	**X**	*optional*
Patient satisfaction[7]	Women who report being satisfied with care using net promoter score (NPS)	Survey	Any/all age groups[8]	**X**	**X**	*optional*
Healthcare worker satisfaction	Healthcare workers (by cadre) [9] who report job satisfaction	Survey	Any/all age groups[8]	**X**	**X**	*optional*
5 pillars of quality of care[10]	Clinical skills, patient-centeredness, health education, data management, and logistics	Survey (using TIP Quality of Care toolkit)	Any/all age groups[8]	**X**	**X**	*optional*
Client volume (population adjusted)	No. of visits (for infant/child wellness checks, child illness, maternal health) No. of ANC visits No. of PNC visits No. of deliveries in facilities	Survey	Any/all age groups[8]	**X**	**X**	*optional*

[1] Several outcomes, including stunting, LBW, and preterm birth, will be tracked as continuous measurements (e.g., weight in grams), but reported as binary outcome variables.

[2] High Hb cut off to be established in accordance with Rwandan protocols. Low Hb levels signify anemia, while high Hb levels have been associated with other MCH risks, including pre-eclampsia and preterm birth.

[3] Stunting outcome is not applicable for measurement in the Prenatal and Newborn Survey to be conducted 12 months after baseline since there will have been insufficient exposure to the OI/E-Heza package to capture stunting outcomes.

[4] Treatment for pre-eclampsia may include magnesium sulfate, antihypertensive medications, monitoring, or other obstetric interventions.

[5] To be defined in accordance with Rwandan protocols. Postpartum hemorrhage is commonly defined as blood loss exceeding 500 mL following vaginal birth and 1,000 mL following cesarean.

[6] For children aged 12-23 months, age-appropriate vaccines are: BCG, three doses of DPT-HepB-Hib (pentavalent), four doses of oral polio vaccine, one dose of IPV, three doses of pneumococcal vaccine, two doses of rotavirus vaccine, and one dose of measles and rubella vaccine.

[7] Patient satisfaction will be measured as part of surveys administered to pregnant women and mothers.

[8] "Any/all age groups" denotes flexibility for outcomes that can be measured in alignment with other measurement intervals as is convenient for the study.

[9] Cadres include nurses and community health workers (CHWs) [10] Per TIP's website, TIP is in the process of validating a quality-of-care evaluation toolkit to consider these 5 pillars.

[10] Definitions and measurement protocols outlined in TIP's toolkit will be used for this outcome measure.

3.3. Cost-effectiveness estimation

The study will produce estimates of cost-effectiveness. Cost effectiveness allows us to objectively compare different interventions on the basis of how much they deliver given their cost. By calculating the cost-effectiveness of OI/E-Heza, we will be able to situate it within the broader set of interventions that look to improve maternal and child health.

We will calculate an incremental cost-effectiveness ratio (ICER), which quantifies the cost per unit effect. In other words, we will derive the additional cost of implementing OI/E-Heza for the gain induced in an outcome by OI/E-Heza; for instance, cost per case of stunting averted, per case of low birthweight (LBW) averted and for each additional ANC visit induced. To do this, we will use estimates of the impact on outcomes of interest derived from the evaluation along with measures of cost. We will use an ingredients-based approach to collect cost data, capturing major categories of cost including materials, equipment, management, communications, transportation, and training. Finer, categorized cost data will enable us to more accurately model potential scenarios where OI/E-Heza is scaled up.

It is worth noting that the OI/E-Heza package aims to increase demand for and access to services as well as improve service quality. Consequently, along with the materials, equipment, management, communications, transportation, and training costs associated directly with OI/E-Heza, treatment sites may experience an increase in cost associated with a higher volume of clients and/or improved level of care provided per client. This may be challenging to capture but being aware of it gives us some chance at either collecting data that measures this indirect increase in cost and/or modeling these costs.

3.4. Measurement

3.4.1. Types of data sources

The study relies on two data sources: (1) direct measurement and (2) health facility and CHW record data. Direct measurement will include anthropometric and clinical measurement for women's hemoglobin levels and children's height and weight, as well as surveys administered to pregnant and postpartum women and to healthcare workers. This will be done across treatment and control sites to ensure consistency and comparability of data. To the extent possible, we will also aim to obtain and analyze data collected by health facilities and CHWs as a verification mechanism. In the treatment sites, the E-Heza data stream will be leveraged for generating additional insights, including dynamics and velocity of change in outcomes. Health facility/CHW and E-Heza data will also be used to measure mortality outcomes at a population level. We discuss the specific details for mortality measurement below and in **Annex 3.**

3.4.2. Timing

Annex 1 shows the evaluation timeline. After obtaining necessary research ethics approvals in month 1, we will randomize health facilities to treatment and control arms. Following study site randomization, we assume a 6-month roll-out period for the OI/E-Heza package in selected treatment sites. Once the OI/E-Heza package is fully implemented, the evaluation will consist of the following surveys:

1) **Baseline survey**: data collection with mother-child dyads in each impact age group to establish pre-intervention levels for outcomes of interest in each study site catchment area. We note that the timing of the baseline survey assumes that the ramping up of the OI/E-Heza package in selected treatment sites does not actively modify outcomes till it is fully deployed at month 8.

2) **Prenatal and Newborn Survey:** a survey conducted 12 months after baseline would assess outcomes among pregnant women and infants for which the OI/E-Heza package may be able to make meaningful changes within one year of implementation (see **Table 1**).

3) **Stunting Survey:** To appropriately measure stunting, the final survey should take place no less than 36 months after implementation begins. This will permit the evaluation to capture the full effects of the OI/E-Heza package during the entire period of early pregnancy to 24-27 months of age. Conducting endline earlier could underestimate the real impact of the OI/E-Heza intervention.

One important consideration for the proposed surveys is their timing: all three surveys should take place at approximately the same time of year to control for the effects of seasonality. In a cross-sectional design, seasonality introduces significant variability which could skew results even if the treatment and control groups are affected equally by seasonality. For instance, the effects of seasonality could overshadow the results of the intervention, making it difficult to detect impacts; or, it may make results appear particularly strong, for example, if implemented in proximity to a lean season as compared to a harvest season.

3.5. Sample sizes and minimum detectable impacts (MDIs)

TIP's experience suggests there are approximately 50 births per month at each health facility, equivalent to 150 births every three months. This number of children born every three months gives us an estimate of the number of living children within each three-month impact age group in facility catchment areas. Thus, at any given time, there should be approximately 150 children per age group in each cluster who are eligible for the study. With that in mind, the sample sizes presented below should be feasible because all of them require fewer than 150 mother-baby dyads per age group per cluster.

We present seven different outcomes as options around which to power the study– stunting, LBW, enrollment in antenatal care (ANC) during the first trimester, completion of four or more ANC visits, infant mortality, neonatal mortality, and prenatal Hb (**Tables 2a – 2e**). For each outcome, we identified sample sizes in terms of number of clusters and number of mother-baby dyads per cluster.

All calculations shown are for a two-arm RCT with 70 study sites (35 per arm). **Annex 2 Tables 2.1** and **2.2** present alternative scenarios based on a total of 90 study sites (45 per arm) and 120 sites (60 per arm). Sample sizes assume 80 percent statistical power, a 0.05 level of statistical significance, and were calculated using an intra-class correlation coefficient (ICC) of 0.00089. The ICC is a statistic that measures how similar outcomes are likely to be within clusters and ICC was calculated at the level of health centers using TIP data on recent mid- and upper-arm circumference (MUAC) values for children under 5. The calculated ICC value is comparable to that found in other studies, e.g., Huang et al. 2023 find a value of 0.0027.

Note on Sample Size. To determine the appropriate sample size for data collection, we calculated MDIs—the smallest impact that can be statistically distinguished from zero—for various sample sizes. The MDI is the smallest change in an outcome that the intervention would have to cause in order to be able to detect it as a statistically significant result. In other words, the study can detect any change in the outcome that is equal to or greater than the MDI. The study will not demonstrate statistically significant results – even if the intervention has an impact – if the effect size is smaller than the MDI.

MDIs are directly related to sample size: the larger the sample size, the smaller an effect that can be detected. In other words, with more observations, we greatly increase our ability to detect smaller effect sizes (smaller MDIs). One way to do this is to increase the number of observations surveyed within each cluster i.e., increase the number of pregnant women and mother-child dyads surveyed within each health facility catchment. The other way to do this is to increase the number of clusters included in the study i.e., increase the number of health facility catchments.

Generally, in clustered RCTs the number of clusters tends to drive the size of the impact that can be detected. However, our study environment seems to have a low ICC based on available information which makes it difficult to recommend a clear pathway to improving our ability to detect small MDIs. However, if expanding the total sample size is something desired (to enable detecting smaller MDIs) and feasible, we would err on the side of expanding the number of health facilities included in the study. This would deliver a larger sample size while protecting against a scenario where the real ICC is potentially higher. It is worth emphasizing that this requires careful consideration of costs and the policy environment: the cost of intervening in more health facilities and the cost of interviewing more respondents in each health facility catchment must be within the study budget; and, (2) the GOR grant permission to include a large number of clusters (larger than 70 clusters) in the study for part or the entire duration of study.

3.5.1. Stunting

Table 2a shows MDIs for stunting prevalence separately for a range of baseline values that we may find in our study sample: the average among districts in the Western Region (40%)[1] and the average among the poorest tertile districts (47%) (NISR, MOH, and ICF 2021). Each row considers a scenario based on a different number of mother-baby dyads per site.

On the lower end, 5 dyads per cluster, or 350 dyads total across the 70 study sites would enable the evaluation to detect a 14.1 - 14.4 percentage point change in stunting. On the higher end, with 25 dyads per site, or 1,750 dyads total, the study could detect a change in stunting of 6.4 - 6.5 percentage points.

Table 2a. *MDIs and sample sizes for stunting*

# of dyads per cluster	Full study sample: total # of dyads across all clusters	MDIs - Stunting	
		(Where baseline rate is 40%)	*(Where baseline rate is 47%)*
5	350	14.1	14.4
10	700	10.0	10.2
15	1,050	8.2	8.4
20	1,400	7.1	7.3
25	1,750	6.4	6.5
75	5,250	3.8	3.8
125	8,750	3.0	3.0
Age group: dyads with children 24- 27 mos.			

[1] According to the Rwanda DHS 2019-20, stunting prevalence values are very similar in the Western and Northern regions: 40.2% and 40.5%, respectively; therefore, we used the lower of the two as the assumed baseline option. To provide a point of comparison for a scenario in which selected districts have higher baseline stunting prevalence, we refer to the statistic provided in Dr. Edgar's concept note for the poorest tertile districts.

In both cases, these MDI values exceed the range of reasonable impacts to expect from stunting reduction programs in LMICs. A systematic review of 14 programs from 19 low- and middle-income countries (LMICs) found that programs had a median 3 percentage point average annual reduction in stunting, with a range of 0.6 to 8.4 (Hossain et al. 2017). To detect a **3-percentage point change**, a very large sample size of **125 dyads per site** would be required (equivalent to **8,750 total dyads** across the entire study).

Given the focus on stunting, we recommend the largest study size feasible given population size and cost constraints. However, such a large sample may not be feasible. Therefore, we also present several other outcome options around which to size the study.

3.5.2. Low birthweight

The first alternative outcome we considered is LBW. We calculated MDIs using the Western districts average (6.2%) and Northern districts average (6.8%) as baseline values (NISR, MOH, and ICF 2021). The MDIs shown in **Table 2b** are below the range of reasonable impacts to expect from integrated nutrition programs in LMICs (da Silva Lopes et al. 2017), which would suggest they are achievable results. But, because baseline LBW averages are already quite low, large declines of more than a few percentage points are not possible.

As with our recommendation around stunting, to design the study around LBW, we would recommend the largest study size feasible given population size and cost constraints. Of the options presented below, this would be **25 dyads per site (1,750 total)** which would enable detection of a **3.1 to 3.3. percentage point change** in LBW. This may be ambitious given that a 3-percentage point decline would effectively cut the incidence of LBW in half.

Table 2b. *MDIs and sample sizes for LBW*

# of dyads per cluster	Full study sample: total # of dyads across all clusters	MDIs - LBW	
		Baseline 6.2%	*Baseline 6.8%*
5	350	7.0*	7.3*
10	700	4.9	5.2
15	1,050	4.0	4.2
20	1,400	3.5	3.7
25	1,750	3.1	3.3
75	5,250	1.9	1.9
125	8,750	1.5	1.5

*Indicates MDIs that are not achievable as rates cannot be negative.
Age group: Dyads with children 0-3 mos.

3.5.3. ANC Attendance

Next, we computed MDIs for two related outcomes of ANC attendance (**Table 2c**, next page). The first is enrollment in ANC during the first trimester, for which MDIs were calculated using a national baseline of 58.7%[2] (NISR, MOH, and ICF 2021). The second outcome is completion of four or more ANC visits during

[2] Sub-regional baseline values were not available for 1st trimester ANC enrollment.

pregnancy, which we computed using the national average of 47% as the baseline[3] (NISR, MOH, and ICF 2021). For the different sample sizes we considered, MDIs for the two ANC outcomes are very similar.

While the literature reflects that different interventions may have a wide range of effect sizes on different measures of ANC attendance, we refer to two studies that suggest that between a 5 to 10 percentage point change may be feasible (Mbuagbaw et al. 2015; Wafula et al. 2022). With that in mind we would recommend a study size that would entail an MDI at the mid-point of that range (7.5 percentage points), i.e., **25** dyads per cluster (1,750 in total). This would enable the study to detect any effects at or above a **6.4 percentage point** change in **first trimester ANC enrollment** and a **6.5 percentage point** change in rates of **completing 4 or more ANC visits**.

Table 2c. *MDIs and sample sizes for ANC*

# of dyads per cluster	Full study sample: total # of dyads across all clusters	MDIs	
		1st Trimester ANC *Baseline 58.7%*	ANC Visits ≥ 4 *Baseline 47.2%*
5	350	14.2	14.4
10	700	10.1	10.2
15	1,050	8.2	8.4
20	1,400	7.2	7.3
25	1,750	6.4	6.5
75	5,250	3.8	3.8
125	8,750	3.0	3.0
Age group for 1st trimester ANC enrollment: pregnant women Age group for ANC Visits ≥ 4: dyads with index children aged 0-3 months			

3.5.4. Hemoglobin in pregnancy - anemia

Much of the literature on interventions to reduce anemia in pregnancy reports outcomes as continuous measures of blood hemoglobin (Hb) concentration, rather than as rates of anemia among the sampled population.[4] The mean Hb level for pregnant women in the northern region is 12.74 g/dL (SD: 1.48) and 12.59 g/dL (SD: 1.68) in the western region (calculated from Rwanda DHS 2019-20; NISR, MOH, and ICF 2021). We note that these averages are above the threshold for anemia in pregnancy, 11.0 g/dL.

Table 2d. *MDIs and sample sizes for Hb in pregnancy*

# of pregnant women per cluster	Full study sample: total # of pregnant women across all clusters	MDIs* – Hb in pregnancy	
		Northern region Baseline: 12.74 g/dL (SD: 1.48)	*Western region Baseline: 12.59 g/dL (SD: 1.68)*
5	350	0.43	0.49
10	700	0.30	0.34
15	1,050	0.25	0.28

[3] Sub-regional baseline values were not available for completion of 4 or more ANC visits.

[4] Collecting and reporting on Hb in pregnancy as a continuous measure will also enable the study to assess high Hb levels, which have also been linked to suboptimal pregnancy and birth outcomes.

20	1,400	0.22	0.24
25	1,750	0.19	0.22
75	5,250	0.11	0.13
125	8,750	0.09	0.10
Age group: women in pregnancy *MDIs are presented in units of g/dL			

The literature on anemia in pregnancy offers a wide range of effect sizes. We refer to two studies that used community health workers and e-health interventions to reduce increase maternal Hb levels by 0.11 g/dL and 0.64 g/dL (Singh et al. 2020; Ilboudo et al. 2021). If we assume that the OI/E-Heza intervention may achieve an effect size somewhere between these two values, a midway point would be 0.38 g/dL.[5]

With a sample size of **10 pregnant women per cluster (700 total)**, the study is powered to detect a change in mean Hb levels of between 0.30 – 0.34 g/dL, which is below that midpoint value. A more conservative estimate, benchmarking against Singh et al. 2020, would require **125 pregnant women per cluster (8,750 total)** to detect a 0.09 – 0.10 g/dL change.

3.5.5. Infant and Neonatal Mortality

Lastly, we calculated MDIs for infant mortality rates (IMR) and neonatal mortality rates (NMR), measured as a ratio of number of deaths among children born to the number of live births. IMR uses deaths up to a year from birth, while NMR is defined as deaths occurring anytime from birth to the first 28 days of life. The average IMR in Rwanda's western districts is 32 infant deaths/1,000 births and in its northern districts is 40 infant deaths/1,000 births (NISR, MOH, and ICF, 2021). Similarly, NMR is 17/1,000 births and 24/1,000 births in northern and western respectively.

Because mortality will require a much larger sample size than would be feasible to capture through a household survey, these outcomes will be assessed in each cluster at a population level using health facility and CHW records and measured as a ratio of number of deaths among children born during the study to the number of live births during the course of the study. **Annex 3** provides more detail.

Table 2e. *MDIs and sample sizes for infant mortality and neonatal mortality*

Time period	Survey for outcome measurement	# of live births per cluster	Full study sample: total # of births	MDIs			
				Northern Region	*Southern Region*	*Northern Region*	*Southern Region*
				IMR *Baseline 3.2%*	**IMR** *Baseline 4.0%*	**NMR** *Baseline 1.7%*	**NMR** *Baseline 2.4%*
12 months	Prenatal & newborn survey	600	42,000	0.6	0.6	0.4	0.5
24 months	Stunting Survey	1,200	84,000	0.5	0.5	0.3	0.4
34 months	Stunting Survey	1,750	119,000	N/A*	N/A*	0.3	0.4

[5] For comparison, mean Hb among women who are anemic is between 10.12– 10.41 g/dL in the two regions, meaning that an average increase of 0.88-0.59 g/dL would be required for these women to no longer be categorized as anemic. All MDIs presented are below this range.

*This study does not allow sufficient time for IMR to be measured over a 34-month period and sample of 119,000 births because birth counts for the denominator will stop 12 months prior to the final Stunting Survey.

For NMR, using health facility and CHW records, the study will count all children born up to one month prior to the survey. In the first row of **Table 2e**, we assume 50 births per month occur in each cluster, which is 600 total births over a 12-month period prior to the Prenatal and Newborn Survey. Multiplied across 70 study clusters, the study could expect to see 42,000 births in total (the denominator in IMR and NMR). The numerator will be a count of any deaths among that group of children that occur during their first 28 days of life. Stopping the birth count one month prior to the survey will allow the study to observe whether a death occurs during the 28 days period after the last recorded birth. For IMR - deaths during the first year of life - the denominator will be all births recorded up to 12 months prior to the survey, and the numerator will be any deaths among those children that occur during their first year of life.

Though any infant mortality is concerning, baseline IMR and NMR values are numerically quite small. This means that reasonable MDIs are also small, which necessitate larger sample sizes for the study to detect changes. Since we do not anticipate a significant additional level of effort or cost to collect mortality data via facility/CHW records, we recommend tracking both outcomes over the longest possible period.

For IMR, 24 months of births (totaling approximately 84,000 births), would allow for detection of **a 0.5 percentage point change** in mortality. **For NMR, 34 months of births (totaling approximately 119,000 births)**, would allow the study to detect a **0.3-0.4 percentage point change**. These effect sizes may also be quite ambitious for the OI/E-Heza intervention to achieve, given the complexity of infant and neonatal mortality, and the numerically very low baseline levels, especially for NMR.

3.5.6. Summary

Table 3 summarizes the evaluation's power to detect impacts at different sample sizes for each of the outcomes we considered (other than NMR and IMR, which are measured differently).

Table 3. *Summary of Sample Size Implications for Study Power by Outcome (70 clusters)*

Sample (# of pregnant women & dyads)	Outcome MDIs are within range in literature				
	Stunting	LBW	1st Trimester ANC	4+ ANC Visits	Hb in pregnancy
5 per cluster / 350 total	No	No	No	No	No
10 per cluster / 700 total	No	No	No	No	Yes
20 per cluster / 1,400 total	No	No	Yes	Yes	Yes
25 per cluster / 1,750 total	No	Maybe*	Yes	Yes	Yes
75 per cluster/5,250	Almost	Maybe*	Yes	Yes	Yes
125 per cluster / 8,750 total	Yes	Yes	Yes	Yes	Yes

*Although the magnitude of the MDI is within the range found in the literature, baseline prevalence of LBW is lower than many other contexts, making further reductions hard to achieve.

- For any study with a sample size smaller than 125 pregnant women/dyads per cluster (8,750 in total across the full study), it is unlikely that the evaluation will be able to detect significant changes in stunting or LBW outcomes:
    - The reductions in stunting that a study with 70 total sites could detect (between 6 – 14 percentage points) may be overly ambitious. A much larger sample size (125 dyads per impact age group per site, or 8,750 dyads across the full study) would be required to detect a

3-percentage point change in stunting, which the literature has shown to be a realistic achievement for effective programming.

    - LBW may similarly be challenging to detect because baseline LBW rates are low, and each percentage point change represents a large proportion of cases.

- We presented five other outcomes: first trimester ANC attendance, completion of 4 or more ANC visits, Hb levels in pregnancy, and infant and neonatal mortality, for which we could power the study if it is not feasible to expand the number of study sites.
- A sample size of 25 pregnant women/dyads per cluster (1,750 total) would enable enough power for detection of ambitious but potentially achievable results, including detection of both ANC outcomes, Hb in pregnancy, and possibly LBW. This is also a feasible sample size given the average population size in each health facility catchment area.
- Given the average birth rate in the clusters, the evaluation may also be able to detect statistically significant changes in mortality rates if births and deaths are consistently tracked from the beginning of the study for a total of 24 months for IMR and 34 months for NMR (see **Annex 3**), though the MDIs for these mortality outcomes translate to at least a 10 percent reduction which is still quite ambitious.

4. Summary and indicative data collection costs

The GOR is invested in improving the health of Rwandan mothers and children, and E-Heza offers a compelling opportunity to deliver an optimized set of interventions both effectively and efficiently. A rigorous and well-designed RCT will quantify and build an understanding of the impact of this OI/E-Heza. In turn, this information will enable policymakers to make evidence-based decisions to address maternal and child health using an optimized intervention package that includes the use of E-Heza.

We suggested a two-arm RCT with a control arm (i.e., business-as-usual) and a full OI/E-Heza treatment arm (i.e., E-Heza enabled health facilities and CHWs). Next, we described a comprehensive set of maternal and child health outcomes, along with health system performance outcomes. Using estimates from the impact evaluation and cost data, we proposed a cost-effectiveness analysis of OI/E-Heza allowing it to be compared to other health interventions and rationalizing possible future investment in it.

We suggested a **46-month evaluation timeline**, to accommodate study set up, implementation, data collection and analysis. After an 8-month initial period to set up the study, which includes 6 months for E-Heza to ramp up in treatment facilities, we propose a **36-month timeframe over which OI/E-Heza is implemented** and data collection activities are carried out, followed by 3 months for final analysis, reporting, and dissemination. This 46-month period will permit the OI/E-Heza package to take effect during children's first two years of life, including the prenatal period.

Our analysis suggested a large sample size is required to detect a 3-percentage point change to stunting. This change in stunting is the median of effect sizes shown in the literature and may be very difficult to achieve. Therefore, we suggested that ANC and prenatal Hb may be a more useful set of outcomes to consider designing the study around, offering more reasonable MDI that are within ranges seen in the literature.

Based on the evaluation design we have described, we calculated an indicative budget for data collection activities. The data collection costs we estimate should be treated as suggesting an order-of-magnitude and should not be seen as accurate estimates; a tendering process will yield true data collection costs for an evaluation of this nature. Additionally, these cost estimates do not include any cost of evaluation and survey instrument design, data analysis or reporting and dissemination.

Data collection is composed of two separate types of activities. First, the choice to use a cross-sectional data collection approach requires discovering the relevant sample each time data collection is conducted i.e., there is a need to do a census (or near-census) to find enough relevant respondents to meet sample needs. As shown in **Figure 2**, there are three times when data are collected (Baseline survey, Prenatal and Newborn Survey, and Stunting Survey), therefore a census will be needed at each of those times. Our calculations suggest that the data collection team will need to approach approximately 1,250 households per health center catchment in order to identify the needed sample of eligible respondents within each health center catchment. The second data collection activity is household interviews targeting the three major respondent groups i.e., pregnant women, 0–3-month-olds and 24–27-month-olds, also three times. A description of key parameters and how they were derived is presented in **Annex 4**.

Cost per successfully completed census interaction and household interview (including any physical measurements such as anthropometrics) was derived from past data collection proposal budgets. Household data collection in Rwanda is expensive, driven by high logistical costs (including transport), labor, management, and taxes. Based on past data, we assumed a cost per census interaction of $1.25 and a cost per household interview ranging from $140 to $175. The higher household interview cost comes from inclusion of height measurement (to establish presence of stunting) of 24–27-month-olds, since this outcome requires two highly trained staff be present together to measure each child.

Table 4*. Indicative costs of data collection*

Census cost	$328,125
Data collection cost	$1,323,000
Total cost	$1,651,125

A user-friendly and customizable cost-model has been provided along with this document for readers to simulate different data collection and cost scenarios, the key results from which are presented in **Table 4**. There may be opportunities to cut data collection costs. For instance, CHWs may have reliable records and knowledge of their communities, which may reduce the extent of the census required to discover the required sample. Additionally, carefully planned logistics can also yield significant reductions to survey costs.

References

Hossain, Muttaquina, Nuzhat Choudhury, Khaleda Adib Binte Abdullah, Prasenjit Mondal, Alan A. Jackson, Judd Walson, and Tahmeed Ahmed. 2017. "Evidence-Based Approaches to Childhood Stunting in Low and Middle Income Countries: A Systematic Review." *Archives of Disease in Childhood* 102(10): 903–9. doi:10.1136/archdischild-2016-311050.

Huang, Shan, Michael Toole, Andre MN Renzaho, Sengchanh Kounnavong, Jennifer J. Watts, and Ben Coghlan. 2023. "Protocol for Integrated Solutions for Healthy Birth, Growth and Development: A Cluster-Randomised Controlled Trial to Evaluate the Effectiveness of a Mixed Nutrition Intervention Package in Reducing Child Undernutrition in Lao People's Democratic Republic." *BMJ Open* 13(8): e066014. doi:10.1136/bmjopen-2022-066014.

Ilboudo, Bernard, Léon G. B. Savadogo, Isidore Traoré, Clément Z. Meda, Maurice Kinda, Issiaka Sombié, Michèle Dramaix-Wilmet, and Philippe Donnen. 2021. "Effect of Personalized Support at Home on the Prevalence of Anemia in Pregnancy in Burkina Faso: A Cluster Randomized Trial." *The American Journal of Tropical Medicine and Hygiene* 105(1): 207–16. doi:10.4269/ajtmh.20-1043.

Mbuagbaw, Lawrence, Nancy Medley, Andrea J Darzi, Marty Richardson, Kesso Habiba Garga, and Pierre Ongolo-Zogo. 2015. "Health System and Community Level Interventions for Improving Antenatal Care Coverage and Health Outcomes." *The Cochrane Database of Systematic Reviews* 2015(12): CD010994. doi:10.1002/14651858.CD010994.pub2.

Menon, Purnima, Rahul Rawat, and Marie Ruel. 2013. "Bringing Rigor to Evaluations of Large-Scale Programs to Improve Infant and Young Child Feeding and Nutrition: The Evaluation Designs for the Alive & Thrive Initiative." *Food and Nutrition Bulletin* 34(3_suppl2): S195–211. doi:10.1177/15648265130343S206.

NISR, MOH, and ICF. 2021. *Rwanda Demographic and Health Survey 2019-20*. https://dhsprogram.com/publications/publication-FR370-DHS-Final-Reports.cfm.

da Silva Lopes, Katharina, Erika Ota, Prakash Shakya, Amarjargal Dagvadorj, Olukunmi Omobolanle Balogun, Juan Pablo Peña-Rosas, Luz Maria De-Regil, and Rintaro Mori. 2017. "Effects of Nutrition Interventions during Pregnancy on Low Birth Weight: An Overview of Systematic Reviews." *BMJ global health* 2(3): e000389. doi:10.1136/bmjgh-2017-000389.

Singh, Jitendra Kumar, Dilaram Acharya, Rajan Paudel, Salila Gautam, Mandira Adhikari, Shambhu Prasad Kushwaha, Ji-Hyuk Park, Seok-Ju Yoo, and Kwan Lee. 2020. "Effects of Female Community Health Volunteer Capacity Building and Text Messaging Intervention on Gestational Weight Gain and Hemoglobin Change Among Pregnant Women in Southern Nepal: A Cluster Randomized Controlled Trial." *Frontiers in Public Health* 8: 312. doi:10.3389/fpubh.2020.00312.

Wafula, Solomon T, Aisha Nalugya, Rornald M Kananura, Richard K Mugambe, Moses Kyangwa, John B Isunju, Betty Kyobe, et al. 2022. "Effect of Community-Level Intervention on Antenatal Care Attendance: A Quasi-Experimental Study among Postpartum Women in Eastern Uganda." *Global Health Action* 15(1): 2141312. doi:10.1080/16549716.2022.2141312.

Annex 1. Evaluation event timing

Month	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46
Start-up and randomization																																															
Research ethics clearance																																														
Defining study sites																																													
Randomization																																													
Survey instrument development and testing																																													
E-Heza Roll Out																																													
E-Heza: planning																																													
E-Heza: ramp up*																																													
E-Heza: fully implemented																																													
Measurement Intervals																																													
Baseline census																																													
Baseline survey**																																													
Prenatal & newborn census																																													
Prenatal & newborn survey																																													
Stunting census																																													
Stunting Survey																																													
Reporting																																													
Analysis: baseline report																																													
Analysis: midline report																																													
Analysis: endline report																																													

*E-Heza ramp up occurs after randomization
**Baseline survey occurs once E-Heza is fully operational

Annex 2. Power and Sample Size Calculations with Alternative Numbers of Clusters

Power and Sample Size Calculations

Acronyms:
ICC: intra-class correlation coefficient
MDI: minimum detectable impact
LBW: low birthweight
ANC: antenatal care
Hb: hemoglobin

Annex Table 2.1. *Sample size calculations assuming 90 total clusters (45 sites per study arm)*

			MDIs							
Dyads per cluster	**Total # of dyads**	**Baseline**	**Stunting**		**LBW**		**1st trimester ANC**	**4+ ANC visits**	**Hb in pregnancy** in g/dL	
			40%	*47%*	*6.2%*	*6.8%*	*58.7%*	*47.2%*	*12.74 (SD: 1.48)*	*12.59 (SD: 1.68)*
5	450		12.4	12.7	6.1	6.4	12.5	12.7	0.38	0.43
10	900		8.8	9.0	4.3	4.5	8.9	9.0	0.27	0.30
15	1,350		7.2	7.3	3.5	3.7	7.2	7.3	0.22	0.25
20	1,800		6.3	6.4	3.1	3.2	6.3	6.4	0.19	0.22
25	2,250		5.6	5.7	2.8	2.9	5.6	5.7	0.17	0.19
70	8,400		3.0	3.0	1.5	1.5	3.0	3.0	0.10	0.12

Note: MDIs for stunting, LBW, 1st trimester ANC enrollment, and 4+ ANC visits are in units of percentage points. MDIs for Hb in pregnancy are in units of g/dL.

Annex Table 2.2. *Sample size calculations assuming 120 total clusters (60 sites per study arm)*

			MDIs							
Dyads per cluster	**Total # of dyads**	**Baseline**	**Stunting**		**LBW**		**1st trimester ANC**	**4+ ANC visits**	**Hb in pregnancy** in g/dL	
			40%	*47%*	*6.2%*	*6.8%*	*58.7%*	*47.2%*	*12.74 (SD: 1.48)*	*12.59 (SD: 1.68)*
5	450		10.7	10.9	5.3	5.5	10.8	10.9	0.32	0.37
10	900		7.6	7.8	3.7	3.9	7.6	7.8	0.23	0.26
15	1,350		6.2	6.3	3.1	3.2	6.3	6.3	0.19	0.21
20	1,800		5.4	5.5	2.7	2.8	5.4	5.5	0.16	0.19
25	2,250		4.8	4.9	2.4	2.5	4.9	4.9	0.15	0.17
70	8,400		3.0	3.0	1.3	1.3	3.0	3.0	0.09	0.10

Note: MDIs for stunting, LBW, 1st trimester ANC enrollment, and 4+ ANC visits are in units of percentage points. MDIs for Hb in pregnancy are in units of g/dL.

Assumptions	
ICC	0.00089
Statistical power	0.80
Level of significance	0.05
R-squared (individual)	0.10
R-squared (group)	0.10

Annex 3. Mortality Measurement Schedule

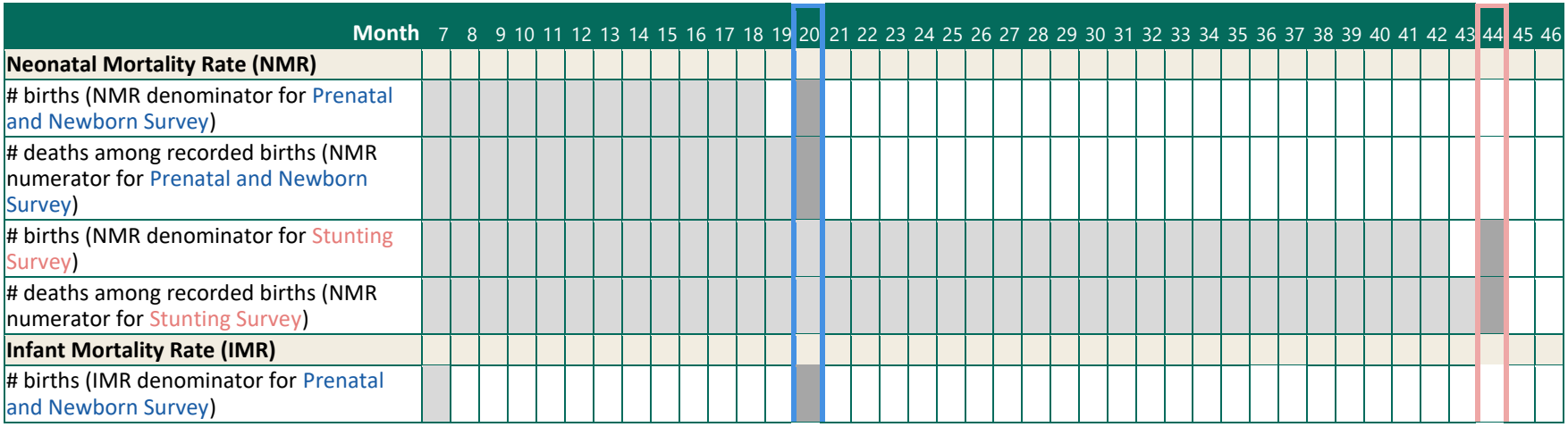
Annex Table 3.1. Mortality Outcomes, Definitions, and Time Lags for Measuring Cases

Outcome	Definitions	Lag between last enrolled birth / pregnancy and study survey
Neonatal mortality rate (NMR)	Infant deaths within birth and 28 days	1 month
Infant mortality rate (IMR)	Infant deaths within birth and 12 months	12 months
Maternal mortality rate (MMR)	Maternal deaths during pregnancy and up to 42 days post-partum	1.5 months
Still birth	Fetal deaths after 28 weeks' gestation	9 months

The four mortality outcomes are measured as ratios. For a death to count toward the numerator of a given ratio, it must occur during the time period specified in the outcome definition. This means there must be a lag between when the last eligible birth is counted toward the denominator and the time that the outcome will be calculated. For example, to calculate neonatal mortality, the study will include births up to one month before the survey date to be able to capture any deaths that may occur among the last enrolled children within 28 days. Similarly, for infant mortality, the study will only count eligible births up to 12 months before the survey date.

The table below reflects the months during which births and deaths are eligible to be counted toward either the numerator or denominator of each outcome, for that outcome to be calculated at either the Prenatal and Newborn Survey or the Stunting Survey.

Annex Table 3.2. Mortality Measurement Sequencing



# deaths among recorded births (IMR numerator for Prenatal and Newborn Survey)	
# births (IMR denominator for Stunting Survey)	
# deaths among recorded births (IMR numerator for Stunting Survey)	
Maternal Mortality Rate (MMR)	
# pregnancies (MMR denominator for Prenatal and Newborn Survey)	
# deaths among women: pregnancy - 42 days postpartum (MMR denominator for Prenatal and Newborn Survey)	
# pregnancies (MMR denominator for Stunting Survey)	
# deaths among women: pregnancy - 42 days postpartum (MMR denominator for Stunting Survey)	
Still birth	
# pregnancies (still birth denominator for Prenatal and Newborn Survey)	
# maternal deaths (still birth denominator for Prenatal and Newborn Survey)	
# pregnancies (still birth denominator for Stunting Survey)	
# maternal deaths (still birth denominator for Stunting Survey)	

Annex 4. Key Parameters in the Cost Model

Cost per interview

Household surveys in Rwanda are expensive. Information from data collection proposals suggest a cost of about $130-180 per completed interview. We assumed that a more involved interview which has anthropometric measurements (e.g., to establish child stunting), will cost $175 while a relatively less involved interview will cost $140. It is worth noting that of the total cost of survey work, about a third of cost is transport and logistics, another third is fees and taxes, and the remainder is labor. Additionally, based on a budget made available to us, 12 enumerators over 22 days (so a total of 264 person days) can interview 600 households. This suggests that each enumerator interviews 2.3 households per day.

Cost per census interaction

Before we are able to interview members of the community, we will need to construct a sample frame – what we call a census activity. For this, we assume a cost per census interaction of $1.25. This is derived from a detailed budget which suggested that the total cost of discovering respondents is about one-quarter the total cost of conducting interviews. Consequently, the cost model is "calibrated" so that total cost of discovering respondents is about one-quarter that of the total cost of interviewing study respondents, which resulted in a per-interaction census cost of $1.25. This cost assumes that each enumerator has 7 census interactions per day.

Number of households to census

While the number of mother-child dyads we intend to interview depends on the sample size calculations, discovering them requires some understanding of local demographics. Using Rwanda DHS data (NISR, MOH, and ICF, 2021), we estimated that 4% of all households have a pregnant woman, 2% have a child between 0-3 months old and 2% have a child 24-27 months old. Our study design suggests a requirement of 10 pregnant women per cluster, 25 0–3-month-olds and 25 24–27-month-olds. Thus, doing a sample building activity where 1,250 households are interviewed in each health center catchment should yield at least 10 pregnant women (it will yield 50 pregnant women), 25 0–3-month-olds and 25 24–27-month-olds.