Transparency for Development: Pre-Analysis Plan

Matthew Bombyk, Jessica Creighton, Akshay Dixit, Dan Levy, Lindsey Roots

Version 2 – April 2018

1 INTRODUCTION

Note: *The Transparency for Development (T4D) program included Randomized Controlled Trials in Indonesia and Tanzania. The analysis for Indonesia and Tanzania will be done separately, treating them as two independent but related projects. This document describes the analysis plan for both projects. A previous version of the pre-analysis plan (pertaining solely to Indonesia) was registered and posted online in November 2017, prior to data collection in Indonesia. This version (Version 2) was updated to include the Tanzania analysis, and was registered prior to data collection in Tanzania.*

The goal of the analysis described here is to identify the effects of the Transparency for Development program intervention on a range of maternal health and community participation outcomes as well as intermediate or process outcomes.

The plan pre-specifies the analysis that will be conducted, before comparing outcomes between treatment and control groups. It outlines the intervention, evaluation design, data sources, hypotheses and outcomes of interest, and the impact estimation strategy.

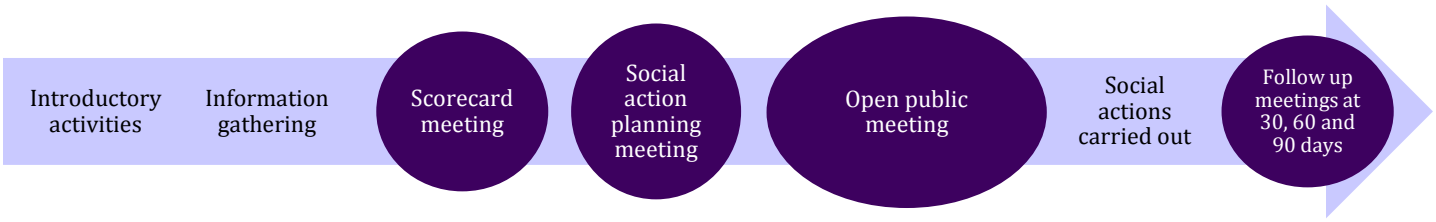
By committing to pre-specified analysis plans we hope to minimize issues of data mining and specification searching. The pre-analysis plan serves the dual purpose of ensuring the endline data collection tools are sufficient for the planned analysis. This plan was written and submitted after baseline data collection and the implementation of the intervention, but prior to the start of endline data collection.

2 OVERVIEW OF THE T4D INTERVENTION

The T4D intervention aims to improve village-level Maternal and Newborn Health (MNH) in rural communities using a modified version of a "community scorecard." The intervention is comprised of seven main activities: (1) introductory activities; (2) information gathering; (3) identification of intervention participants; (4) facilitation of meetings to share information and develop an action plan; (5) sharing the action plan with the greater community during a public meeting; (6) community-led social action; and (7) a series of facilitated follow-up meetings. T4D partnered with two civil society organizations (CSOs) to administer the intervention. The CSO partners are PATTIRO in Indonesia and the local chapter of the Clinton Health Access Initiative (CHAI) in Tanzania.

The intervention components are described briefly below and illustrated in Figure 1. A comprehensive description can be found in the *T4D Intervention Design Report.*[1]

Figure 1. Intervention Activities



2.1 Introductory Activities

At the start of the intervention, CSO-employed facilitators entered assigned villages and began meeting with village leadership, community health volunteers[2], and citizens. At these meetings, the facilitators explained the intervention and its aims, identified survey respondents, collected MNH data, and identified potential intervention participants, or "community activists." This introduction was also intended to encourage ownership of the project by community members.

2.2 Information Gathering

Scorecard data was collected using two types of surveys: the first was a beneficiary survey administered to 20 – 30 women in each village. These women had given birth in the two years prior to the intervention. This survey included questions on interactions between the women and the health system, and the take-up of key MNH services. The second survey was a simple facility survey to collect data on infrastructure, cleanliness, and human resource availability.

2.3 Identification of Participants

Fifteen to sixteen community members in each village were selected to formally participate in the intervention. Known as "community activists" or CAs, these participants were recruited based on a number of characteristics, including: personal interest in maternal and neonatal health, time and willingness to volunteer, and enthusiasm about improving the village. Formal leaders and health

[1] T4D, "Transparency for Development Intervention Design," April 2016, http://www.t4dev.org/sites/default/files/file-uploads/Intervention%20Design%20Description%202016%2004_0.pdf.

[2] Indonesia only

workers were excluded.

2.4 COMMUNITY ACTIVIST MEETINGS

The facilitator worked intensively with the community activists over a two-day period, with the goal of formulating a comprehensive action plan to improve MNH in the village. The first day focused on sharing the collected data in the form of a community scorecard; the second day was spent developing a social action plan to address problems revealed by the information in the scorecard.

2.4.1 Scorecard Meeting

Scorecard information on the uptake of three key MNH health levers was presented to community activists: antenatal care (ANC) (Tanzania only), birth preparedness planning (Indonesia only)[3], delivery (both countries), and post-natal services (PNC) (both countries). The facilitators used the information from the levers to start a discussion about the barriers to improved MNH in the village[4]. Once barriers were identified and discussed, community activists were presented with short vignettes of actions that other communities have taken to improve service delivery, uptake, or both. These vignettes, or "social action stories," served two purposes. First, they were intended to build the confidence of the community activists by illustrating the ability of communities to solve their own problems. Second, they introduced a variety of different ways for communities to try to improve uptake and care, in an effort to stimulate thinking about which might be appropriate to the context of the particular community.

2.4.2 Social Action Planning Meeting

On the second day, facilitators led the community activists through the process of developing a plan of action to improve MNH – the social action plan. Community activists were prompted to formulate a mix of actions, ideally including actions that could lead to improvements within 90 days (the formal duration of the intervention) and those that could lead to improvements over the longer term.

---

[3] ANC uptake in Indonesia is high, so the lever instead focused on birth preparedness planning.

[4] Examples of barriers include: transportation to the health facility, knowledge of proper health seeking behavior, and mistreatment by facility staff.

2.5 Open Public Meeting

After the community activists developed the social action plan, an open public meeting was held to share an abbreviated version of the community scorecard and the social action plan. Comments and additional suggestions were solicited, and other community members were invited to voluntarily participate in future actions.

2.6 Social Actions

After these initial meetings, the community activists were expected to carry out the actions developed in the social action plan. The community activists worked on these actions independently, without the help of CSO facilitators. Example social actions included organizing a community education campaign on the importance of ANC during pregnancy and confronting a healthcare provider who had been stealing medicine.

2.7 Follow-up Meetings

The CSO facilitator convened three follow-up meetings with the community activist group. These meetings occurred approximately every 30 days, allowing the facilitator to check in with the community activists on the progress made on the social actions and to discuss revisions, new actions and, ultimately, a sustainability plan.

Figure 2. T4D MNH Health Levers

Tanzania:

- Antenatal care within the first 12 weeks and 4 total visits
- Delivery at a health facility with a skilled provider
- Postnatal check-up within 7 days (mother and newborn)

Indonesia:

- Comprehensive birth preparedness plan
- Delivery at a health facility with a skilled provider
- Postnatal check-up within 7 days for the mother and 48 hours for the newborn

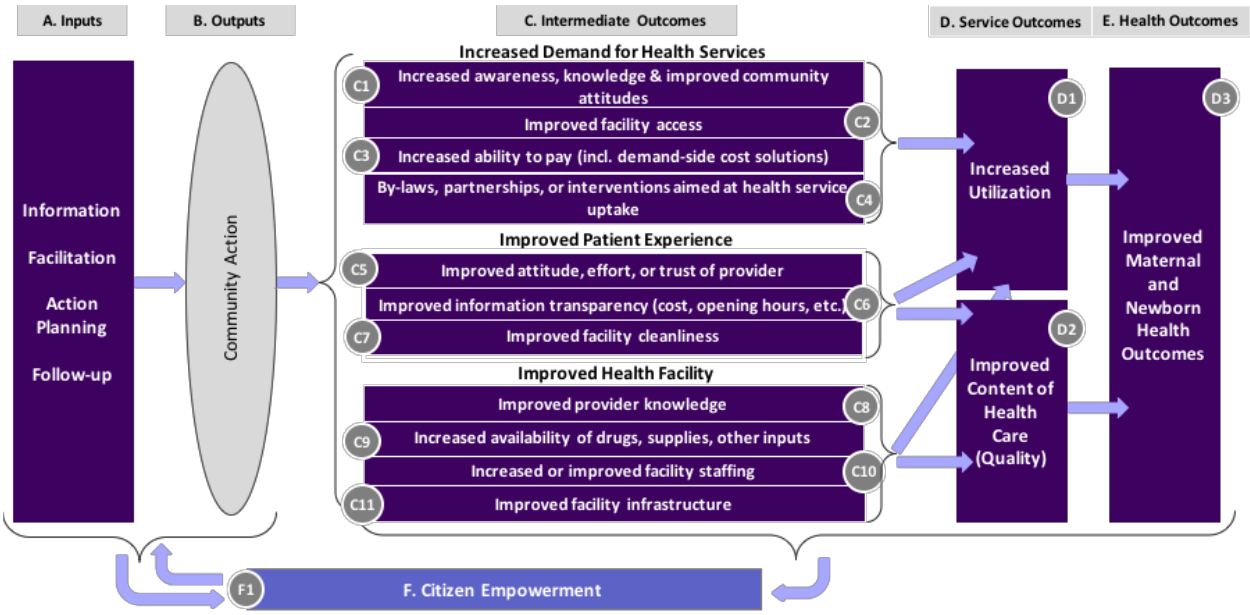
3 PROJECT LOGIC AND RESEARCH QUESTIONS

3.1 LOGIC MODEL

Figure 3 illustrates how the intervention is hypothesized to affect health outcomes. To have an impact, the community must understand and be motivated by the information, develop a plausible social action plan, and successfully carry it out. This process may have an impact and improve health outcomes through three main pathways:

1) The proportion of people receiving services increases (increased utilization);
2) The quality of services delivered through existing channels improves (improved content of clinical care); and
3) People who were receiving lower quality care at one outlet choose to seek care at a higher quality outlet.

Figure 3. Logic Model of the Intervention



This intervention is designed to primarily trigger (1) and (2)—collective action targeted at improving service utilization (D1 in Figure 3), the content of clinical health care (D2 in Figure 3), or both. These pathways form the basis of T4D research questions #1 and #2, described in the next section. Since the information component of the intervention does not inform communities of the relative quality of health facilities, the T4D team does not expect the intervention to explicitly trigger (3), communities seeking care at different outlets.

Community activists may choose to carry out a range of social actions (B in Figure 3). These social actions trigger one or more intermediate outcomes (C in *Figure 3*), such as awareness of activities mothers should undertake during pregnancy, or a change in midwife behavior, which can lead to an impact on utilization of healthcare services, content of healthcare services, or both (D in *Figure 3*). This ultimately improves health outcomes (E in *Figure 3*), including decreasing neonatal and infant mortality.

Since these actions are entirely designed and undertaken by community members, the intervention may also improve citizen participation and sense of empowerment (F in *Figure 3*). This is particularly to the extent that the actions facilitate unfamiliar experiences where community members engage with each other and with providers and public officials in an attempt to diagnose and alleviate problems with a public service that they value (A-B in *Figure 3*). If these actions are then successful in improving that service, they can create a positive feedback loop – participants become aware of their ability to improve their health care, which fosters further empowerment and encourages participation in additional or more sustained efforts to diagnose and alleviate problems, thereby increasing the improvement of community health service and outcomes (C-E in *Figure 3*).

3.2 RESEARCH QUESTIONS

The evaluation will seek to answer the following key research questions, which form the main hypotheses that the study will test:

1) What is the effect of the intervention on the **utilization of health care services** related to maternal and child health?

2) What is the effect of the intervention on the **content of health care** services related to maternal and child health?

3) What is the effect of the intervention on **health outcomes**?

4) What is the effect of the intervention on **citizen empowerment and efficacy**, both perceived and actual?

5) If there are significant effects, what are the **mechanisms** through which these effects occur?

6) What is the role of **context** in shaping or determining these mechanisms?

Research questions 1-4 will be assessed through the use of randomized controlled trials (RCTs), while research question 5 will be explored through the tracking of community actions and analysis of intermediate outcomes. Research question 6 will be the subject of sub-group, cross-country, and qualitative analysis. The following sections primarily detail how the impact evaluation will answer research questions 1-4.

3.3 Levels of Analysis – Primary, Secondary and Intermediate

The analysis of project impact is composed of two levels of analysis:

1) We specify a **primary** set of outcomes and regression models that will serve as the main measure of overall project impact. These specifications were carefully chosen to address all of the key research questions, minimize the likelihood of false claims of impact, and maintain the ability to detect any project impacts, if they occurred. Positive results from these analyses can be used to make judgments on overall project impact.
2) An analysis of the **secondary and intermediate outcomes** will seek to provide some understanding of the mechanisms through which the project had an impact (if it had one), and the contextual factors that affected the impact. If the project did not have an impact on the primary outcomes, the analysis of secondary and intermediate outcomes might shed some light on why it did not by exploring where the causal chain is likely to have been broken. Also included in this tier is the **sub-group analyses**. For these analyses, we will not be as concerned about Type I statistical errors, as we will not be using them to make judgments on the project's overall impact. Instead, they will be used to better understand any impacts that are found in the primary analysis.

We will perform the same primary analysis for Indonesia and Tanzania separately, treating them as two independent but related projects, which mirrors the reality of the project implementation. Since a major goal of this evaluation is to discover why any differences in project impact occurred, contrasting the results from the two countries will be much more fruitful than pooling the results and attempting to look at an "average impact" of the project across countries.

4 IMPACT EVALUATION DESIGN

The design of this impact evaluation study relies on randomly assigning a set of villages to be the target of project activities. This section describes the specifics of the randomization procedure, sampling methods, and timing of data collection.

4.1 TREATMENT ASSIGNMENT

A Randomized Controlled Trial (RCT) will be used to asses the impact of the interventions. By randomly assigning communities to treatment and control groups, RCTs ensure that the two groups are equivalent at the outset of the intervention. If well designed and implemented, this method ensures that any differences in outcomes between the two groups that are observed after the intervention are due to the intervention and not to other factors.

The impact evaluation design consists of two-armed RCTs in both Indonesia and Tanzania, with 100 treatment and 100 control villages in each country. In Indonesia, the study villages are split between two provinces, with 85 villages in Banten and 115 in South Sulawesi. In Tanzania, the villages are split between two regions, with 77 villages in Dodoma and 123 in Tanga.

4.2 UNIT OF RANDOMIZATION

The unit of randomization is the health facility. Health facilities generally serve several villages, and sometimes these overlap. To deal with this, a procedure was developed to sample a single village (or sometimes two, in the case of Tanzania) per health facility. The health facility was then randomly assigned to treatment or control groups (along with its associated village or villages). As a result:

- In Indonesia, the 200 sample villages map one-to-one onto 200 facilities.
- In Tanzania, the 200 villages map many-to-one onto 153 facilities, with no more than two villages associated with a single facility.

The project implementation occurred at the village level, only in the sampled villages(s) associated with treatment health facilities.

4.3 STRATIFICATION

Random assignment of villages to treatment and control groups was stratified on a few key variables in both countries. The sample of villages in each case was divided into strata, and then villages within each stratum were assigned to either treatment or control.

In Tanzania, the T4D team chose to stratify by region, proportion of women in the village who have delivered in a health facility, and whether there are one or two sample villages in the catchment area of the health facility. Since stratification involves dividing the sample up into distinct "buckets", stratification variables need to be discrete. While region and the number of sample villages in the catchment area of the facility are both discrete, the proportion of women who have delivered in a health facility is not. In order to stratify on this variable, the team generated a dummy for whether or not the proportion of women in the village who have delivered in a facility was above or below the sample median, and then stratified on this variable instead.

In Indonesia, the T4D team chose to stratify by province and the proportion of women in the village who had delivered in a health facility. Since each sample village in Indonesia corresponded to a unique health facility, the additional stratification variable used in Tanzania was unnecessary.

4.4 SAMPLING ACROSS SURVEY ROUNDS

The study follows a repeated cross-section design, whereby independent samples of households are selected from each village in baseline and follow up. This was because the key household respondents are women who have recently had a child, so interviewing the same women at baseline and endline might not yield data of a recent pregnancy on both occasions. Instead, households with women who had given birth in the 12 months prior to the survey will be interviewed, at both baseline and endline. Thus, household-level changes are not possible to measure, and only village-level outcomes at baseline can be used as controls. One advantage of this study design is that sample attrition is not a concern.

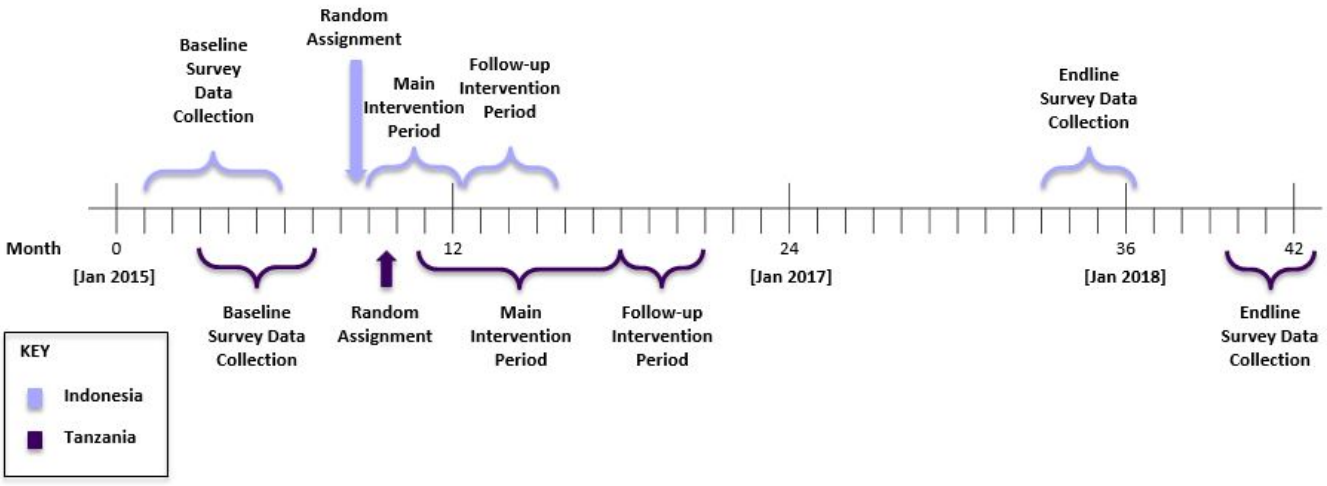
At baseline, T4D conducted interviews with a total of 5,398 household respondents (3,000 in Tanzania and 2,398 in Indonesia). The team anticipates conducting about 12,000 household interviews at endline (6,000 in each country), as a larger sample will enable more precise estimates of impact.

4.5 TIMING OF DATA COLLECTION

In Indonesia, baseline data collection took place from February to June 2015. Following data collection and random assignment, the intervention was rolled out in two waves. The main intervention period consisted of the community scorecard, social action planning, and open meetings, and the start of the social actions. The follow up intervention period involved the continuation of the social actions and all follow-up meetings. Endline survey data collection will commence approximately 21 months after the completion of the main intervention period, beginning in November 2017. The period between the end of the main intervention and the start of data collection allows time for the conception and birth of a new cohort of infants.

In Tanzania, baseline data collection took place from March to July 2015. Instead of two waves, the Tanzania intervention was rolled out in four waves, meaning the main intervention period lasted longer than it did in Indonesia. As a result, endline data collection will begin about 6 months later in Tanzania than in Indonesia, with an anticipated start date in May 2018.

Figure 4 – Impact Evaluation Timeline



4.6 BASELINE EQUIVALENCE

In both countries, we verified that the treatment and control groups do indeed look similar on a host of baseline characteristics. Overall, the differences between the groups tend to be very small in magnitude and rarely statistically significant. For Indonesia, only five out of the 96 baseline

variables (including key outcomes) turned out to generate a treatment coefficient that was statistically significant at the 5% level[5], which falls within the expected bounds of naturally occurring sample variation and pure chance. Similarly, for Tanzania, only six of 112 variables tested generated a statistically significant treatment coefficient at the 5% level[6]. For details, see Chapter 4.4 of the Baseline Report.

5 IMPACT ESTIMATION STRATEGY

5.1 REGRESSION MODEL

Given the use of random assignment to select treatment sites, the basic method of estimating program impacts consists of comparing mean outcomes for the treatment and control groups. The estimation strategy consists of estimating the following regression equation:

(1) $$Y_{ijk} = \beta 0 \quad + \beta_1 TREAT_{jk} + \beta_2 \boldsymbol{STRATA}_k + \varepsilon_{ijk}$$

In this equation, the variable Y_{ijk} is the outcome of interest (whether the mother gave birth at a birth facility, weight-for-age of child, etc.) for mother/child *i* in village *j* in catchment area *k*. The variable $\boldsymbol{STRATA}_k$ is a vector of dummy variables that indicate the randomization strata[7]. The variable $TREAT_{jk}$ is an indicator variable that takes a value of one if the village was assigned to receive the treatment, and zero otherwise. The coefficient β_1 provides the estimate of the impact of the program. Standard errors will be clustered at the facility level, which is the level of treatment assignment.

[5] The variables were: 1) ANC check - mother received urine sample results; 2) woman ever had an ANC visit because of a complication; 3) proportion of women paying for post-natal care; 4) in most recent effort government officials/political leaders listened to, and took seriously their proposal; 5) in past year, respondent or anyone in the household has participated in an information or election campaign.

[6] The variables were: 1) whether or not anyone in the household owns a bicycle; 2) the satisfaction dummy for
whether or not the respondent felt they were properly informed of what was happening during recent visit to the health facility; 3) whether the respondent gave birth in a private hospital; 4) type of transport taken to facility for delivery (bicycle); 5) type of transport taken to facility for delivery (public transportation); 6) proportion underweight (weight-for age).

[7] Random treatment assignment of health facilities was stratified by region, proportion of women in the village who have delivered in a health facility and (in Tanzania) whether there are one or two sample villages in the catchment area of the health facility.

Given that T4D collected baseline data on households in the 200 villages and plans to collect endline data on a different set of households within these same villages, the team also plans to estimate a second set of regressions that control for the village-level average of the relevant outcome variable at baseline (Y_j^0) and of other village characteristics (X_j^0) at baseline (Equation 2)[8]. These additional explanatory variables are meant to increase the statistical precision of the impact estimates, but the research team does not expect it to have a substantial effect on the magnitude of the impact estimates (because of random assignment to treatment and control groups).

(2) $$Y_{ijk} = \beta_0 + \beta_1 TREAT_{jk} + \beta_2 \boldsymbol{STRATA}_k + \beta_3 Y_j^0 + \beta_4 X_j^0 + \varepsilon_{ijk}$$

The coefficients for β_1 from Equations (1) and (2) will be compared, and if there is no substantial difference in the point estimates, Equation (2) will be used as the primary specification. If there is a difference, Equation (1) will be used as the primary specification. Results from both specifications will be reported.

5.2 Adjustment of p-values for Multiple Inference

The T4D team is measuring impact separately on two distinct "families" of primary outcomes – health and healthcare (listed in sections 7.1.1, 7.1.2, 7.1.3, 8.1.1, 8.1.2 and 8.1.3), and empowerment (listed in section 7.1.4 and 8.1.4). Both these families have multiple outcomes, and there is a chance of over-rejection of the null hypothesis of no impact owing to multiple hypothesis testing. At a 95% level of confidence, one could expect to find an impact on one out of 20 variables purely due to chance. Checks against multiple hypothesis testing include ex-post adjustment of p-values, and reducing the number of hypotheses tested ex-ante.

The approach we will take for the primary outcomes within each "family" is to control the False Discovery Rate (FDR), which limits the expected proportion of rejections within a hypothesis that are Type I errors (Benjamini, Krieger, and Yekutieli 2006[9]; Anderson 2008[10]; Casey, Glennerster,

---

[8] Some variables were not measured in the baseline. In these cases, the baseline control is either omitted, or a similar proxy variable is used.

[9] Yoav Benjamini, Abba M. Krieger, and Daniel Yekutieli, "Adaptive linear step-up procedures that control the false discovery rate", *Biometrika*, 93, 3 (2006): *pp.* 491-507

[10] Michael L. Anderson, "Multiple Inference and Gender Differences in the Effects of Early Intervention: A Reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects," *Journal of the American Statistical Association* 103, no. 484 (December 1, 2008): 1481–95.

and Miguel 2012[11]). This will guard against false rejections of the null hypothesis for key outcomes and therefore against falsely declaring statistically significant the overall impacts of the project. We will not adjust p-values for secondary or intermediate outcomes, since these will be used to improve our understanding of the project's impact[12], and not to judge the success of the project. Since the primary outcome analysis will be performed separately for Tanzania and Indonesia, this multiple hypothesis adjustment will be applied separately for each country.

Adjusting p-values for multiple hypothesis testing reduces the power to detect effects for each outcome individually, and hence it is important to limit the number of outcomes considered. To that end, certain hypotheses will be grouped into indices. The construction of indices is described in more detail in section 9.

6 SUB-GROUP ANALYSIS

The T4D researchers plan to estimate the impacts of the program on a number of key sub-groups.

First, the team will examine geographic variability within countries, by separately considering the project impact for the two regions (in Tanzania), or two provinces (in Indonesia).

Second, for all outcomes, the team will look at villages associated with three sub-groups of facilities: those with 1) high, 2) medium, and 3) low quality of health services at baseline, as these may potentially affect the perceived value of the health system to intervention participants and the efficacy of any improvements or increased utilization on ultimate health outcomes. Quality of health services at baseline will be measured using baseline data on facility infrastructure/supplies, and community perceptions of facility quality.

The team will also look at village level characteristics that potentially affect the village's willingness and ability to act collectively. Specifically, the team will look at baseline data on community level demographics, perceptions of the responsiveness of providers and other public officials[13], and

[11] Katherine Casey, Rachel Glennerster, and Edward Miguel, "Reshaping Institutions: Evidence on Aid Impacts Using a Preanalysis Plan," *The Quarterly Journal of Economics* 127, no. 4 (November 1, 2012): 1755–1812.

[12] This is discussed in Section 3.3.

[13] The team expects participation to differ according to perceptions of provider and public official responsiveness, as implied by the Five Worlds Framework; Kosack and Fung, "Does Transparency Improve Governance?"

measures of trust and solidarity (e.g. willingness of community members to commit time and or money to communal activities) and collective action (e.g. rates of participation in communal activities). These will be used to create an index, which will then be used to assign each village a "collective action capacity score" of High, Medium, or Low.

Finally, the intervention targets health outcomes around the time of birth, yet the endline data collection will be based on a sample of mothers who gave birth in the 12 months prior to the survey date. Hence, the team will also explore whether the impact of the program on health outcomes (see Section 7.1.3) is different for mothers who gave birth closer to the date of the survey (i.e. between 0-6 months) than those who gave birth at a later date (i.e. between 6-12 months).

Being explicit about the sub-groups at this stage is important to protect the research against conducting statistical tests *ex-post* and discovering spurious results. While the T4D team does not wish to discard the possibility of testing hypotheses that emerge from the implementation of the project and the qualitative work, the team will be explicit about which hypotheses were specified at the outset and which ones arose after the design work.

To conduct sub-group analyses, we will use the regression strategy described in section 5.1 but adding dummy variables for the sub-groups and an interaction between the treatment dummy variable and the relevant sub-group. The coefficient on the interaction will represent the difference in the impact of the program for that sub-group relative to the omitted sub-group.

7 OUTCOMES OF INTEREST – INDONESIA

This section describes the various outcomes that will be used in the quantitative analysis of the T4D project in Indonesia. These are split into three groups:

- Primary outcomes, which will be used to make a judgment on the overall impact of the project
- Secondary outcomes, which are important final outcomes but will not be used to make a judgment about project impact
- Intermediate outcomes, which will be analyzed to uncover the mechanisms through which the primary and secondary outcomes were impacted by the project

7.1 PRIMARY OUTCOMES

This section contains a description of the outcomes the T4D project will use to assess the impact of the intervention, beginning with the primary outcomes. The tables below provide the list of primary outcomes organized by research question, along with the definition and the key justification for including each outcome.

7.1.1 Research Question 1: Uptake of Health Services

The outcome measures used to measure Research Question 1 link directly to the health levers presented to community activities during the intervention scorecard meeting (see Figure 2).

Outcome	Definition	Key Justification
Delivery with a skilled birth attendant	Whether the respondent delivered with a skilled birth attendant.	Birth in a facility with a skilled attendant is one of the three information levers in the T4D intervention in Tanzania and Indonesia. A skilled attendant at birth is another one of the 11 cores health indicators monitored by the UN commission on Accountability for Women's and Children's Health.[14] Appropriate medical attention during delivery is linked to reduction in complications that can cause serious illness or death to the mother and newborn,[15] and thus can contribute to reduction in neonatal and infant mortality rates, as well as the maternal mortality ratio.
Delivery at a health facility	Whether the respondent delivered at a health facility.	Birth in a facility with a skilled attendant is one of the three information levers in the T4D intervention in Tanzania and Indonesia. Birth in a facility ensures a sanitary environment and easier access to emergency services should complications arise. The Australia Indonesia Partnership for Maternal and Neonatal Health (AIPMNH) found the risk of death for infants to be six times higher if a birth occurs at home with a TBA instead of at a health facility.[16]

[14] "Recommendation 2: Health Indicators."

[15] Statistics Indonesia et al., "Indonesia Demographic and Health Survey 2012."

[16] A Abdullah et al., "Maternal Health & Risk Factors Associated with Neonatal Death in AIPMNH-Assisted Districts in NTT: A Matched Case-Control Study" (Indonesia: Australia Indonesia Partnership for Maternal and Neonatal Health, September 2014), http://aipmnh.org/web_en/images/reports/Book_Case_Control_Study_Risk_Factor_Neonatal_Deaths_FINAL_June_2015.pdf.

Post-partum care (mother) & Post-natal care (newborn)	Postpartum care – Whether the respondent received at least one post-partum check with a skilled attendant, within 7 days of giving birth. Postnatal care – Whether the newborn received at least one post-natal check with a skilled attendant, within 7 days of birth. Postpartum and postnatal care utilization will be combined and treated as one outcome. They are both binary variables, and will be combined by creating a single binary variable on whether a respondent received **both** post-partum and post-natal care.	Early post-partum/post-natal care for mothers and babies is one of the three information levers in the T4D intervention in Tanzania and Indonesia. Post-natal care for mothers and babies within two days of birth is one of 11 core health indicators monitored by the UN Commission on Accountability for Women's and Children's Health.[17] Additionally, the WHO recommends that for a facility birth, mothers and newborns should receive post-natal care in the facility for at least 24 hours after birth. For home births, a post-natal visit should occur within 24 hours after birth and "at least 3 additional post-natal contacts are recommended for all mothers and newborns, on day 3 (48-72 hours), between days 7-14 after birth and six weeks after birth."[18]

[17] "Recommendation 2: Health Indicators."

[18] "WHO Recommendations on Postnatal Care of the Mother and Newborn" (Geneva, Switzerland, 2013), http://www.who.int/maternal_child_adolescent/documents/postnatal-care-recommendations/en/.

7.1.2 Research Question 2: Content of Health Services

Delivery content of care will be an integer variable with values ranging from 0 to 6, calculated as the sum of a set of 6 binary variables corresponding to the components described in the table below.

Postpartum content of care will be an integer variable with values ranging from 0 to 9, calculated as the sum of 9 binary variables corresponding to the components described in the table below.

Postnatal content of care will be an integer variable with values ranging from 0 to 9, calculated as the sum of 9 binary variables corresponding to the components described in the table below.

These three outcomes – delivery, postpartum, and postnatal content of care – will be combined into an index (see Section 9 for details on how this index will be constructed), and treated as **one single outcome on content of care**.

Component	**Definition**	**Key Justification**
Delivery content of care	Number of delivery content of care components received by the respondent: 1. Initiation of breastfeeding within one hour of birth 2. Baby wrapped to mother (skin-to-skin contact) within 30 minutes of birth 3. Delayed bathing for 6 hours - deliveries at home only 4. [Oxytocin] injection right after delivery (after the baby but before the placenta) 5. Uterine massage after delivery of placenta - applicable for vaginal delivery only 6. Clean and dry cord care	Quality of care associated with one of the three information levers in the T4D intervention in Indonesia and Tanzania.
Postpartum content of care (mother)	Number of postpartum content of care components received by the respondent: Physical checks of the mother to treat complications that arise from delivery 1. Blood pressure 2. Checked breasts 3. Check for bleeding 4. Examine perineum	Quality of care associated with one of the three information levers in the T4D intervention in Indonesia and Tanzania.

	Advice to mothers on how to care for themselves and their children 5. Danger signs for newborns 6. Danger signs for mothers 7. Breastfeeding 8. Family planning/contraception Provision of vitamins and supplements 9. Vitamin A	
Postnatal content of care (newborn)	Number of postnatal content of care components received by the infant: Physical checks of the newborn to treat complications that arise from delivery 1. Baby weighed 2. Body examined for danger signs "generally examined/looked at baby's body" 3. Checked cord Provision of recommended vaccines 4. Polio 5. Hepatitis B (HB0) 6. BCG 7. DPT-HB Provision of recommended vitamins or supplements 8. Vitamin K1 9. Eye cream	Quality of care associated with one of the three information levers in the T4D intervention in Indonesia and Tanzania.

7.1.3 Research Question 3: Health Outcomes[19]

Outcome	Definition	Key Justification
Weight-for-age	Weight-for-age z-score. Whether the infant is below 2 standard deviations from the median WHO Child Growth Standards.	Weight-for-age is a measure of chronic and acute malnutrition.[20] In principle, better antenatal care, including the provision of micronutrient supplements, nutritional advice, and the treatment of maternal illness could increase infant height and weight, as could vaccinations and growth monitoring.[21] Studies of similar interventions have shown significant effects on this measure.[22]

[19] Health outcomes for MNH typically center on mortality rates and ratios. Based on the focus of the intervention, T4D anticipates the health outcomes most likely to be effected are maternal mortality, infant mortality, neonatal mortality, birth weight and weight-for-age. The T4D project is not powered to detect changes in the maternal mortality ratio, and will instead use literature to link birth in a facility and skilled birth attendance to maternal mortality. Additionally, since the baseline confirmed that not all babies are weighed at birth (especially in Tanzania), the study is not set up to measure birth weight directly, and there is likely a difference between those infants who are weighed at birth and those who are not, the T4D project will not look at birth weight as a primary health outcome. There is evidence, however, of a correlation between birth weight and weig-for-age (Tanzania National Bureau of Statistics and ICF Macro, "Tanzania Demographic and Health Survey 2010," 164), which will be assessed.

[20] Tanzania National Bureau of Statistics and ICF Macro, "Tanzania Demographic and Health Survey 2010," 162.

[21] Paul J. Gertler and Christel Vermeersch, "Using Performance Incentives to Improve Health Outcomes," World Bank Policy Research Working Paper (Rochester, NY: The World Bank, June 1, 2012).

[22] Martina Björkman and Jakob Svensson, "Power to the People: Evidence from a Randomized Field Experiment on Community-Based Monitoring in Uganda," The Quarterly Journal of Economics 124, no. 2 (May 1, 2009): 735–69; Gertler and Vermeersch, "Using Performance Incentives to Improve Health Outcomes."

Height-for-age	Height-for-age z-score. Whether the infant is below 2 standard deviations from the median WHO Child Growth Standards.[23]	Height-for-age is a measure of chronic malnutrition. Stunting is affected by both chronic and recurrent illness and, unlike weight-indicators, is not sensitive to recent, short-term changes to diet. In principle, better antenatal care, including the provision of micronutrient supplements, nutritional advice, and the treatment of maternal illness could increase infant height and weight, as could vaccinations and growth monitoring.[24] Dodoma region in Tanzania (one of the T4D intervention areas) is one of 4 regions in Tanzania where stunting exceeds 50%.[25] Stunting is also a specific area of concern in Indonesia.

[23] "Child Growth Standards: Weight-for-Age."

[24] Gertler and Vermeersch, "Using Performance Incentives to Improve Health Outcomes."

[25] Tanzania National Bureau of Statistics and ICF Macro, "Tanzania Demographic and Health Survey 2010," 162–63.

7.1.4 Research Question 4: Empowerment

Outcome	Definition	Key Justification
Participation	Index of activities associated with empowerment and efficacy. The following three outcomes are binary variables, and will be combined into an index following the procedure described in Section 9. 1) Whether the respondent reported that she (or a household member) participated in communal activities over the previous 12 months, in which people came together to work for the benefit of the community. 2) Whether the respondent reported that over the previous 12 months, people in her neighborhood or village had gotten together to petition government officials or political leaders for something benefiting the community. 3) Whether the respondent reported that she (or a household member) had done at least one of the following in the past 12 months – • attended a village or neighborhood council meeting, public hearing, discussion group • met with a politician, called him/her, or sent a letter • participated in a protest or demonstration • participated in an information or election campaign • alerted newspaper, radio or TV to a local problem • notified police or court about a local problem • online activism (such as reporting problems on Facebook, Twitter, etc.)	All intervention outcomes stem from community action. The intervention's efficacy may differ according to how familiar participants are with similar community actions, and the experience of participants and those in their network with the intervention may make them more or less likely to participate in similar actions in the future. The T4D team will assess 3 types of community action to cover the range of types of participation that may result from the intervention: communal self-help activities, communal appeals to officials (a proxy for "long route" actions in the T4D intervention), and individual participation in a range of public-facing political and social actions.

Perceptions of empowerment	The perception of the respondent about her power to make important decisions and take actions that improve life in her village, for herself and others. This will be assessed on a 4-point scale, where 1 means being totally unable to improve life in this village, and 4 means having full control to make important decisions and actions to improve life in this village. Response bias related to differing understanding of empowerment between respondents will be removed by comparing responses to respondents' understanding of three "vignettes" describing individuals of the same gender as the respondent trying to improve the teaching at their local school, with varying degrees of success. Respondents' understanding of the levels of empowerment of the individuals in these vignettes will be modeled as a function of the respondent's age, educational level, an index of assets, and answers to the three participation questions above. Thresholds in perceptions for each respondent will be adjusted for subjective biases that vary systematically across these groups, allowing comparable results across individuals and communities.	In principle, intervention participants may perceive greater empowerment, particularly if their actions lead to noticeable improvements in the quality or responsiveness of health and health care. Because empowerment perceptions are inherently subjective and can differ systematically across groups—particularly marginalized groups that may rationalize or not recognize their disempowerment—the team will use anchoring vignettes[26] which can correct for group-level subjective biases among respondents (see Masset (2015) for an application specifically to empowerment).[27]

[26] Gary King et al., "Enhancing the Validity and Cross-Cultural Comparability of Measurement in Survey Research," American Political Science Review 98 (2004): 191–207.

[27] Edoardo Masset, "Measuring Empowerment in Rural India Using Vignettes," Journal of Development Effectiveness 7, no. 3 (July 3, 2015): 346–56.

7.2 SECONDARY OUTCOMES

In addition to the primary outcomes outlined above, the T4D team intends to measure secondary outcomes, which are important final outcomes but will not be used to make a judgment about project impact. These outcomes pertain to the first three Research Questions.

Three of the secondary outcomes pertain to ANC. While the Indonesia intervention levers don't target ANC specifically, ANC is intimately linked to MNH. Birth weight is also included because it is an outcome that is closely linked to the uptake and quality of ANC. Another secondary outcome, birth preparedness planning, is one of the Indonesia levers, but the T4D team considers it secondary because it is a tool to encourage the primary outcomes facility delivery and skilled care at birth.

Lastly, the time period around pregnancy is one when women are most likely to have a major depressive episode, and mothers' depression is associated with adverse developmental outcomes for children. While mental health is not explicitly part of the T4D logic model, there may be a link between the T4D intervention and respondents' perceptions of control over pregnancy and delivery, and since literature suggests a link between perception of control and depression, the impact of the T4D intervention on maternal depression is a hypothesis the team wishes to explore further.[28]

The secondary outcomes are listed below.

Research question	Outcome	Definition	Key Justification
1. Uptake of health Services	Four or more ANC visits	Whether the respondent attended four or more antenatal care visits	Antenatal care coverage is one of 11 core health indicators monitored by

[28] Atif Rahman et al., "Impact of Maternal Depression on Infant Nutritional Status and Illness: A Cohort Study," Archives of General Psychiatry 61, no. 9 (September 1, 2004): 946–52; Jenn Leiferman, "The Effect of Maternal Depressive Symptomatology on Maternal Behaviors Associated with Child Health," Health Education & Behavior 29, no. 5 (October 2002): 596–607; E. Mark Cummings and Patrick T. Davies, "Maternal Depression and Child Development," Journal of Child Psychology and Psychiatry 35, no. 1 (January 1, 1994): 73–122; S. R. Cogill et al., "Impact Of Maternal Postnatal Depression On Cognitive Development Of Young Children," British Medical Journal (Clinical Research Edition) 292, no. 6529 (1986): 1165–67.

		with a skilled provider.	the UN Commission on Accountability for Women's and Children's Health[29] and is a common indicator used to assess the uptake of ANC. Additionally, the number of ANC visits is linked to delivery in a health facility.[30]
1. Uptake of health services	First ANC visit within the first trimester	Whether the respondent had a first antenatal care visit within the first 13 weeks of pregnancy with a skilled provider.	Early care-seeking behavior is important because there is a positive relationship between ANC care and facility delivery. ANC care also gets expectant mothers into the healthcare system earlier.
1. Uptake of health services	Birth preparedness	Number of birth preparedness activities conducted by the respondent: 1. Where to deliver the baby 2. Who will assist with the birth 3. Transportation to place of delivery 4. Payment for delivery 5. Identification of a compatible blood donor 6. Support to look after children while away (for women with childcare responsibilities only) 7. Support to look after the home	Comprehensive birth preparedness planning is one of the three information levers in the T4D intervention in Indonesia. The WHO recommendations on health promotion interventions for maternal and newborn health 2015 include a "strong recommendation" for birth preparedness and complication readiness.[31] The specific components include those outlined in the official "labor planning and prevention of complications" sticker provided with the 2016 edition of the Indonesia

[29] "Accountability for Women's and Children's Health: Recommendation 2: Health Indicators," *WHO*, 2016, http://www.who.int/woman_child_accountability/progress_information/recommendation2/en/.

[30] Statistics Indonesia et al., "Indonesia Demographic and Health Survey 2012."

[31] "WHO Recommendations on Health Promotion Interventions for Maternal and Newborn Health 2015" (Geneva, Switzerland: World Health Organization, 2015), http://apps.who.int//iris/bitstream/10665/172427/1/9789241508742_report_eng.pdf?ua=1.

		while away	Department of Health's *Mother and Child Health Book*.[32]
2. Content of health services	Antenatal content of care	Number of antenatal content of care components received by the respondent during one or more antenatal care visits: 1. Iron tablets or syrup 2. Blood pressure measurement 3. Urine sample 4. Tetanus toxoid injection 5. Informed of signs of pregnancy complications 6. Counseled on birth preparedness planning 7. Counseled on nutrition	
3. Health outcomes	Birth weight	Whether the infant has a birth weight less than 2500g	Health outcomes for MNH typically center on mortality rates and ratios. Based on the focus of the intervention, T4D anticipates the health outcomes most likely to be effected are maternal mortality, infant mortality, neonatal mortality, birth weight and weight-for-age. While the study is not set up to measure birth weight directly and there is likely a difference between those infants who are weighed at birth and those who are not, the T4D project will look at birth weight as a secondary health outcome.
3. Health	Maternal	Respondent's score on the Kessler Psychological Distress Scale (K6)[33].	The time around pregnancy is one when women are most likely to have a

[32] "Mother and Child Health Book" (Cilegon City Government Health Service and the Indonesia Department of Health, 2016).

outcomes	depression	This will be measured by converting the K6 to a 0-24 scale (each of the six questions coded 0-4 and summed), with 13+ indicating SMI.	major depressive episode. A mother's depression is highly associated with adverse developmental outcomes for children in social, emotional, and cognitive domains. Since perceptions of control are strongly linked to depression, and the T4D team expects that the intervention may provide women greater control over the circumstances of their pregnancies and deliveries (including ability to access health care during this period), the T4D intervention may lead to lower levels of depression.[34]

[33] "National Comorbidity Survey: K10 and K6 Scales," Harvard Medical School, 2005, http://www.hcp.med.harvard.edu/ncs/k6_scales.php.

[34] Rahman et al., "Impact of Maternal Depression on Infant Nutritional Status and Illness: A Cohort Study"; Leiferman, "The Effect of Maternal Depressive Symptomatology on Maternal Behaviors Associated with Child Health"; Cummings and Davies, "Maternal Depression and Child Development"; Cogill et al., "Impact Of Maternal Postnatal Depression On Cognitive Development Of Young Children."

7.3 INTERMEDIATE OUTCOMES

The T4D team will examine a number of intermediate outcomes, designed to assess the various pathways through which T4D may see impact on the outcomes under Research Questions 1 and 2. These outcomes link directly to the T4D intervention logic model (see Column C in Figure 3), and are associated with the social actions taken by intervention communities. Based on further analysis of qualitative data, the T4D team may amend this list of intermediate outcomes to include any additional pathways of impact that might emerge.

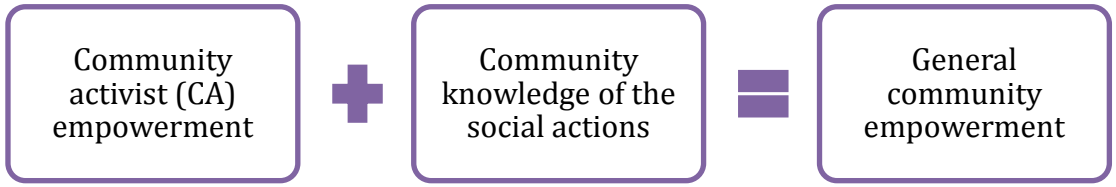
Logic Model Pathway	**Outcome definition**
Increased awareness, knowledge and improved community attitudes	Whether the respondent is able to correctly answer questions on healthcare practices**.**
Improved facility access (transportation, new facility, longer facility hours, outreach services)	**Build or request a new facility** - If (and how many) new facilities have been built/commissioned in the last 3 years.
Improved facility access (transportation, new facility, longer facility hours, outreach services)	**Community organized transportation** – • If the respondent reported using any form of transportation to go to the facility for delivery • If the respondent reported using an ambulance to go the facility for delivery • If the respondent reported using community-organized transportation to go to the facility for delivery
Improved facility access (transportation, new facility, longer facility hours, outreach services)	**Fix road** - Whether the respondent reported the fixing or building of new roads or bridges in the last 3 years, or the procurement of inflatable boats in the last 3 years, for traveling to the health facility.
Improved facility access (transportation, new facility, longer facility hours, outreach services)	**Travel time -** How long it took for the respondent to get to the facility
Improved facility access (transportation, new facility, longer facility hours, outreach services)	**Travel cost -** Amount of money spent by the respondent for transportation to get to the facility

Improved facility access (transportation, new facility, longer facility hours, outreach services)	**Mobile clinic or outreach services** - If (and how many) new mobile maternity clinics have been set-up in the last 3 years.
Improved facility access (transportation, new facility, longer facility hours, outreach services)	**New posyandu** - If a new posyandu had been started, a physical posyandu structure had been built, or the posyandu cadre had been reactivated in the past 3 years
Improved facility access (transportation, new facility, longer facility hours, outreach services)	**Request ambulance** - Whether the facility has a functional ambulance.
Increased ability to pay	**Address cost of services** - Whether the respondent reported cost of care as a barrier to utilization of care, whether the respondent paid a fee for the delivery, total fees for the delivery reported by the respondent, whether the respondent (or spouse) has an insurance/health protection program enrolment
Increased ability to pay	**Raise community funds for delivery or other maternity costs, to support hospital patients (mothers)** - Whether the respondent was part of a woman's savings group to help save for costs associated with MNH care, whether the respondent reported using a community fund to help pay for costs associated with MNH care.
Bylaws, partnerships, or other interventions aimed at health system uptake	**By-laws -** If any (and how many) of the villages have a by-law or other legal measure, requiring/encouraging the uptake of MNH services.
Bylaws, partnerships, or other interventions aimed at health system uptake	**Midwife-TBA partnerships** - If there are midwife partnerships with baby dukun or TBAs
Improved attitude, effort, trust of the provider (includes increased availability of provider)	**Monitor or complain about health facility staff performance** - The level of satisfaction reported by the respondent with the quality of MNH services, whether the respondent reported non-dignified care, provider neglect, or abuse, and provider perception of community feedback and information levels.
Improved attitude, effort, trust of the provider (includes increased availability of provider)	**Midwife residence in village** - If a midwife is i. Assigned to the village ii. Lives in the village, and iii. Receives free housing in the village.
Improved facility cleanliness	Linked to actions around **cleaning the facility, or complaints about the cleanliness of the facility** - Level of cleanliness reported by the respondent, Observed level of cleanliness of the facility delivery room and toilet

Improved information transparency (cost, opening hours, etc.) or complaint mechanisms	**Cost transparency** - If cost information for delivery and other services is displayed publically
Improved information transparency (cost, opening hours, etc.) or complaint mechanisms	**Hours transparency** - If information on facility operating hours is displayed publically.
Improved information transparency (cost, opening hours, etc.) or complaint mechanisms	**Complaint mechanism** - i. If the facility has a complaint management system ii. If the facility conducts routine meetings with community members in the service area to identify how to improve quality of services
Improved provider knowledge	Linked to actions on **educating or training midwives**, this is defined as the number of questions the midwife can correctly answer questions on healthcare practices.
Improved facility infrastructure	**Fixing facility infrastructure** – Electricity, Telecommunications and computer/internet, Water, Delivery room (including privacy and beds), toilet, communication (i.e. how would patients contact the facility).
Increased availability of drugs, supplies and other inputs	**Address shortage of medicines or supplies** - whether the facility has in stock essential medicines/equipment/vaccines for mothers and children, and basic supplies/equipment.
Increased or improved facility staffing	Number of staff in the MNH and birth unit, number of vacancies for staff.

The T4D team will also explore pathways linked to Research Question 4. In order for empowerment to transfer to the general community, the team hypothesizes that there must first be an increase in community activist empowerment, that the community activists must carry out social actions, and finally, that the general community must be aware that these actions took place (see Figure 5).

Figure 5 –Empowerment



The following intermediate outcomes measure community knowledge of the actions.

Logic Model Pathway	**Outcome definition**
Knowledge of social actions aimed at improving health outcomes	Number of social actions reported by respondents as having been carried out by Community Activists
Knowledge of social actions aimed at improving health outcomes	Number of social actions that impacted the respondent positively

8 OUTCOMES OF INTEREST - TANZANIA

This section describes the various outcomes that will be used in the quantitative analysis of the T4D project in Tanzania. As with Indonesia, these are split into three groups:

- Primary outcomes, which will be used to make a judgment on the overall impact of the project
- Secondary outcomes, which are important final outcomes but will not be used to make a judgment about project impact
- Intermediate outcomes, which will be analyzed to uncover the mechanisms through which the primary and secondary outcomes were impacted by the project

These outcomes are largely similar to the ones listed above for Indonesia. There are, however, some noteworthy differences. For the purpose of clarity, all outcomes for Tanzania are listed below – including the ones defined in exactly the same way as Indonesia.

8.1 PRIMARY OUTCOMES

This section contains a description of the outcomes the T4D project will use to assess the impact of the intervention, beginning with the primary outcomes. The tables below provide the list of primary outcomes organized by research question, along with the definition and the key justification for including each outcome.

8.1.1 Research Question 1: Uptake of Health Services

The outcome measures used to measure Research Question 1 link directly to the health levers presented to community activities during the intervention scorecard meeting (see Figure 2).

Outcome	Definition	Key Justification
Four or more ANC visits	Whether the respondent attended four or more antenatal care visits with a skilled provider.	Antenatal care (a minimum of four visits total, with the first visit occurring within the first trimester) is one of the three information levers in the T4D intervention in Tanzania. Antenatal care coverage is one of 11 core health indicators monitored by the UN Commission on Accountability for Women's and Children's Health[35] and is a common indicator used to assess the uptake of ANC. Additionally, the number of ANC visits is linked to delivery in a health facility.[36]
First ANC visit within the first trimester	Whether the respondent had a first antenatal care visit within the first 13 weeks of pregnancy with a skilled provider.	Antenatal care (a minimum of four visits total, with the first visit occurring within the first trimester) is one of the three information levers in the T4D intervention in Tanzania. Early care-seeking behavior is important because there is a positive relationship between ANC care and facility delivery. ANC care also gets expectant mothers into the healthcare system earlier.
Delivery with a skilled birth attendant	Whether the respondent delivered with a skilled birth attendant.	Birth in a facility with a skilled attendant is one of the three information levers in the T4D intervention in Tanzania and Indonesia. A skilled attendant at birth is another one of the 11 cores health indicators monitored by the UN commission on Accountability for Women's and Children's Health.[37] Appropriate medical attention

[35] "Accountability for Women's and Children's Health: Recommendation 2: Health Indicators," *WHO*, 2016, http://www.who.int/woman_child_accountability/progress_information/recommendation2/en/.

[36] Statistics Indonesia et al., "Indonesia Demographic and Health Survey 2012."

[37] "Recommendation 2: Health Indicators."

		during delivery is linked to reduction in complications that can cause serious illness or death to the mother and newborn,[38] and thus can contribute to reduction in neonatal and infant mortality rates, as well as the maternal mortality ratio.
Delivery at a health facility	Whether the respondent delivered at a health facility.	Birth in a facility with a skilled attendant is one of the three information levers in the T4D intervention in Tanzania and Indonesia. Birth in a facility ensures a sanitary environment and easier access to emergency services should complications arise. The Australia Indonesia Partnership for Maternal and Neonatal Health (AIPMNH) found the risk of death for infants to be six times higher if a birth occurs at home with a TBA instead of at a health facility.[39]
Post-partum care (mother) & Post-natal care (newborn)	Postpartum care – Whether the respondent received at least one post-partum check with a skilled attendant, within 7 days of giving birth. Postnatal care – Whether the newborn received at least one post-natal check with a skilled attendant, within 7 days of birth. Postpartum and postnatal care utilization will be combined and treated as one outcome. They are both binary	Early post-partum/post-natal care for mothers and babies is one of the three information levers in the T4D intervention in Tanzania and Indonesia. Post-natal care for mothers and babies within two days of birth is one of 11 core health indicators monitored by the UN Commission on Accountability for Women's and Children's Health.[40] Additionally, the WHO recommends that for a facility birth, mothers and newborns should receive post-natal care in the facility for at least 24 hours after birth. For home births, a post-natal visit should occur within 24 hours after birth and "at least 3 additional post-natal contacts are recommended for all

[38] Statistics Indonesia et al., "Indonesia Demographic and Health Survey 2012."

[39] A Abdullah et al., "Maternal Health & Risk Factors Associated with Neonatal Death in AIPMNH-Assisted Districts in NTT: A Matched Case-Control Study" (Indonesia: Australia Indonesia Partnership for Maternal and Neonatal Health, September 2014), http://aipmnh.org/web_en/images/reports/Book_Case_Control_Study_Risk_Factor_Neonatal_Deaths_FINAL_June_2015.pdf.

[40] "Recommendation 2: Health Indicators."

	variables, and will be combined by creating a single binary variable on whether a respondent received **both** post-partum and post-natal care.	mothers and newborns, on day 3 (48-72 hours), between days 7-14 after birth and six weeks after birth."[41]

8.1.2 Research Question 2: Content of Health Services

Antenatal content of care will be an integer variable with values ranging from 0 to 11, calculated as the sum of a set of 11 binary variables corresponding to the components described in the table below.

Delivery content of care will be an integer variable with values ranging from 0 to 6, calculated as the sum of a set of 6 binary variables corresponding to the components described in the table below.

Postpartum content of care will be an integer variable with values ranging from 0 to 9, calculated as the sum of 9 binary variables corresponding to the components described in the table below.

Postnatal content of care will be an integer variable with values ranging from 0 to 6, calculated as the sum of 6 binary variables corresponding to the components described in the table below.

These four outcomes – antenatal, delivery, postpartum, and postnatal content of care – will be combined into an index (see Section 9 for details on how this index will be constructed), and treated as **one single outcome on content of care**.

Component	**Definition**	**Key Justification**
Antenatal content of care	Number of antenatal content of care components received by the respondent during one or more antenatal care visits: 1. Blood pressure	Quality of care associated with one of the three information levers in the T4D intervention in Tanzania.

[41] "WHO Recommendations on Postnatal Care of the Mother and Newborn" (Geneva, Switzerland, 2013), http://www.who.int/maternal_child_adolescent/documents/postnatal-care-recommendations/en/.

	2. Urine sample 3. Blood sample 4. HIV testing 5. Tetanus toxoid injection 6. Counseling on nutrition 7. Iron tablets/syrup 8. Medication for deworming 9. Medication to prevent malaria 10. Counseled on birth preparedness planning 11. Counseled on nutrition	
Delivery content of care	Number of delivery content of care components received by the respondent: 1. Initiation of breastfeeding within one hour of birth 2. Baby wrapped to mother (skin-to-skin contact) within 30 minutes of birth 3. Delayed bathing for 6 hours - deliveries at home only 4. [Oxytocin] injection right after delivery (after the baby but before the placenta) 5. Uterine massage after delivery of placenta - applicable for vaginal delivery only 6. Clean and dry cord care	Quality of care associated with one of the three information levers in the T4D intervention in Indonesia and Tanzania.
Postpartum content of care (mother)	Number of postpartum content of care components received by the respondent: Physical checks of the mother to treat complications that arise from delivery 1. Blood pressure 2. Checked breasts 3. Check for bleeding 4. Examine perineum Advice to mothers on how to care for themselves and their children 5. Danger signs for newborns 6. Danger signs for mothers 7. Breastfeeding 8. Family planning/contraception	Quality of care associated with one of the three information levers in the T4D intervention in Indonesia and Tanzania.

	Provision of vitamins and supplements 9. Vitamin A	
Postnatal content of care (newborn)	Number of postnatal content of care components received by the infant: Physical checks of the newborn to treat complications that arise from delivery 1. Baby weighed 2. Body examined for danger signs "generally examined/looked at baby's body" 3. Checked cord Provision of recommended vaccines 4. Polio 5. BCG 6. DPT-HB	Quality of care associated with one of the three information levers in the T4D intervention in Indonesia and Tanzania.

8.1.3 Research Question 3: Health Outcomes[42]

Outcome	Definition	Key Justification
Weight-for-age	Weight-for-age z-score. Whether the infant is below 2 standard deviations from the median WHO Child Growth Standards.	Weight-for-age is a measure of chronic and acute malnutrition.[43] In principle, better antenatal care, including the provision of micronutrient supplements, nutritional advice, and the treatment of

[42] Health outcomes for MNH typically center on mortality rates and ratios. Based on the focus of the intervention, T4D anticipates the health outcomes most likely to be effected are maternal mortality, infant mortality, neonatal mortality, birth weight and weight-for-age. The T4D project is not powered to detect changes in the maternal mortality ratio, and will instead use literature to link birth in a facility and skilled birth attendance to maternal mortality. Additionally, since the baseline confirmed that not all babies are weighed at birth (especially in Tanzania), the study is not set up to measure birth weight directly, and there is likely a difference between those infants who are weighed at birth and those who are not, the T4D project will not look at birth weight as a primary health outcome. There is evidence, however, of a correlation between birth weight and weig-for-age (Tanzania National Bureau of Statistics and ICF Macro, "Tanzania Demographic and Health Survey 2010," 164), which will be assessed.

[43] Tanzania National Bureau of Statistics and ICF Macro, "Tanzania Demographic and Health Survey 2010," 162.

		maternal illness could increase infant height and weight, as could vaccinations and growth monitoring.[44] Studies of similar interventions have shown significant effects on this measure.[45]
Height-for-age	Height-for-age z-score. Whether the infant is below 2 standard deviations from the median WHO Child Growth Standards.[46]	Height-for-age is a measure of chronic malnutrition. Stunting is affected by both chronic and recurrent illness and, unlike weight-indicators, is not sensitive to recent, short-term changes to diet. In principle, better antenatal care, including the provision of micronutrient supplements, nutritional advice, and the treatment of maternal illness could increase infant height and weight, as could vaccinations and growth monitoring.[47] Dodoma region in Tanzania (one of the T4D intervention areas) is one of 4 regions in Tanzania where stunting exceeds 50%.[48] Stunting is also a specific area of concern in Indonesia.

[44] Paul J. Gertler and Christel Vermeersch, "Using Performance Incentives to Improve Health Outcomes," World Bank Policy Research Working Paper (Rochester, NY: The World Bank, June 1, 2012).

[45] Martina Björkman and Jakob Svensson, "Power to the People: Evidence from a Randomized Field Experiment on Community-Based Monitoring in Uganda," The Quarterly Journal of Economics 124, no. 2 (May 1, 2009): 735–69; Gertler and Vermeersch, "Using Performance Incentives to Improve Health Outcomes."

[46] "Child Growth Standards: Weight-for-Age."

[47] Gertler and Vermeersch, "Using Performance Incentives to Improve Health Outcomes."

[48] Tanzania National Bureau of Statistics and ICF Macro, "Tanzania Demographic and Health Survey 2010," 162–63.

8.1.4 Research Question 4: Empowerment

Outcome	Definition	Key Justification
Participation	Index of activities associated with empowerment and efficacy. The following three outcomes are binary variables, and will be combined into an index following the procedure described in Section 9. 1) Whether the respondent reported that she (or a household member) participated in communal activities over the previous 12 months, in which people came together to work for the benefit of the community. 2) Whether the respondent reported that over the previous 12 months, people in her neighborhood or village had gotten together to petition government officials or political leaders for something benefiting the community. 3) Whether the respondent reported that she (or a household member) had done at least one of the following in the past 12 months – • attended a village or neighborhood council meeting, public hearing, discussion group • met with a politician, called him/her, or sent a letter • participated in a protest or demonstration • participated in an information or election campaign • alerted newspaper, radio or TV to a local problem • notified police or court about a local problem. • online activism (such as reporting problems on Facebook, Twitter, etc.)	All intervention outcomes stem from community action. The intervention's efficacy may differ according to how familiar participants are with similar community actions, and the experience of participants and those in their network with the intervention may make them more or less likely to participate in similar actions in the future. The T4D team will assess 3 types of community action to cover the range of types of participation that may result from the intervention: communal self-help activities, communal appeals to officials (a proxy for "long route" actions in the T4D intervention), and individual participation in a range of public-facing political and social actions.

Perceptions of empowerment	The perception of the respondent about her power to make important decisions and take actions that improve life in her village, for herself and others. This will be assessed on a 4-point scale, where 1 means being totally unable to improve life in this village, and 4 means having full control to make important decisions and actions to improve life in this village. Response bias related to differing understanding of empowerment between respondents will be removed by comparing responses to respondents' understanding of three "vignettes" describing individuals of the same gender as the respondent trying to improve the teaching at their local school, with varying degrees of success. Respondents' understanding of the levels of empowerment of the individuals in these vignettes will be modeled as a function of the respondent's age, educational level, an index of assets, and answers to the three participation questions above. Thresholds in perceptions for each respondent will be adjusted for subjective biases that vary systematically across these groups, allowing comparable results across individuals and communities.	In principle, intervention participants may perceive greater empowerment, particularly if their actions lead to noticeable improvements in the quality or responsiveness of health and health care. Because empowerment perceptions are inherently subjective and can differ systematically across groups—particularly marginalized groups that may rationalize or not recognize their disempowerment—the team will use anchoring vignettes[49] which can correct for group-level subjective biases among respondents (see Masset (2015) for an application specifically to empowerment).[50]

[49] Gary King et al., "Enhancing the Validity and Cross-Cultural Comparability of Measurement in Survey Research," American Political Science Review 98 (2004): 191–207.

[50] Edoardo Masset, "Measuring Empowerment in Rural India Using Vignettes," Journal of Development Effectiveness 7, no. 3 (July 3, 2015): 346–56.

8.2 SECONDARY OUTCOMES

As with Indonesia, the T4D team intends to measure secondary outcomes, which are important final outcomes but will not be used to make a judgment about project impact. These outcomes pertain to the first and third Research Questions.

Birth weight is included as an outcome because it is closely linked to the uptake and quality of ANC, one of the information levers presented as part of the T4D intervention in Tanzania. It is not included as a primary outcome, because baseline data revealed a low percentage of babies weighed at birth (49.2%). Another secondary outcome, birth preparedness planning, was one of the T4D information levers in Indonesia (but not Tanzania). It is a tool to encourage the primary outcomes facility delivery and skilled care at birth.

Lastly, the time period around pregnancy is one when women are most likely to have a major depressive episode, and mothers' depression is associated with adverse developmental outcomes for children. While mental health is not explicitly part of the T4D logic model, there may be a link between the T4D intervention and respondents' perceptions of control over pregnancy and delivery, and since literature suggests a link between perception of control and depression, the impact of the T4D intervention on maternal depression is a hypothesis the team wishes to explore further.[51]

The secondary outcomes are listed below.

Research question	Outcome	Definition	Key Justification
1. Uptake of health services	Birth preparedness	Number of birth preparedness activities conducted by the respondent:	Comprehensive birth preparedness planning is one of the three information levers in the T4D intervention in Indonesia. The WHO

[51] Atif Rahman et al., "Impact of Maternal Depression on Infant Nutritional Status and Illness: A Cohort Study," Archives of General Psychiatry 61, no. 9 (September 1, 2004): 946–52; Jenn Leiferman, "The Effect of Maternal Depressive Symptomatology on Maternal Behaviors Associated with Child Health," Health Education & Behavior 29, no. 5 (October 2002): 596–607; E. Mark Cummings and Patrick T. Davies, "Maternal Depression and Child Development," Journal of Child Psychology and Psychiatry 35, no. 1 (January 1, 1994): 73–122; S. R. Cogill et al., "Impact Of Maternal Postnatal Depression On Cognitive Development Of Young Children," British Medical Journal (Clinical Research Edition) 292, no. 6529 (1986): 1165–67.

		8. Where to deliver the baby 9. Who will assist with the birth 10. Transportation to place of delivery 11. Payment for delivery 12. Identification of a compatible blood donor 13. Support to look after children while away (for women with childcare responsibilities only) 14. Support to look after the home while away	recommendations on health promotion interventions for maternal and newborn health 2015 include a "strong recommendation" for birth preparedness and complication readiness.[52] The specific components include those outlined in the official "labor planning and prevention of complications" sticker provided with the 2016 edition of the Indonesia Department of Health's *Mother and Child Health Book*.[53]
3. Health outcomes	Birth weight	Whether the infant has a birth weight less than 2500g. Where a recorded measure of birth weight is not available, the respondent (the infant's mother) will be asked about the size of the infant at birth. The responses will be coded into a binary variable corresponding to low birth weight. Further, even where a recorded measure of birth weight is available, the respondent will be asked about the size of the infant at birth. These responses will be used to predict the birth weight status for cases where a recorded measure of birth weight was not available.	Health outcomes for MNH typically center on mortality rates and ratios. Based on the focus of the intervention, T4D anticipates the health outcomes most likely to be effected are maternal mortality, infant mortality, neonatal mortality, birth weight and weight-for-age. While the study is not set up to measure birth weight directly and there is likely a difference between those infants who are weighed at birth and those who are not, the T4D project will look at birth weight as a secondary health outcome.

[52] "WHO Recommendations on Health Promotion Interventions for Maternal and Newborn Health 2015" (Geneva, Switzerland: World Health Organization, 2015), http://apps.who.int//iris/bitstream/10665/172427/1/9789241508742_report_eng.pdf?ua=1.

[53] "Mother and Child Health Book" (Cilegon City Government Health Service and the Indonesia Department of Health, 2016).

		For these cases, an additional dummy variable will be added to the regression specification, indicated whether birth weight was predicted.	

3. Health outcomes	Maternal depression	Respondent's score on the Kessler Psychological Distress Scale (K6)[54]. This will be measured by converting the K6 to a 0-24 scale (each of the six questions coded 0-4 and summed), with 13+ indicating SMI.	The time around pregnancy is one when women are most likely to have a major depressive episode. A mother's depression is highly associated with adverse developmental outcomes for children in social, emotional, and cognitive domains. Since perceptions of control are strongly linked to depression, and the T4D team expects that the intervention may provide women greater control over the circumstances of their pregnancies and deliveries (including ability to access health care during this period), the T4D intervention may lead to lower levels of depression.[55]

[54] "National Comorbidity Survey: K10 and K6 Scales," Harvard Medical School, 2005, http://www.hcp.med.harvard.edu/ncs/k6_scales.php.

[55] Rahman et al., "Impact of Maternal Depression on Infant Nutritional Status and Illness: A Cohort Study"; Leiferman, "The Effect of Maternal Depressive Symptomatology on Maternal Behaviors Associated with Child Health"; Cummings and Davies, "Maternal Depression and Child Development"; Cogill et al., "Impact Of Maternal Postnatal Depression On Cognitive Development Of Young Children."

8.3 INTERMEDIATE OUTCOMES

The T4D team will examine a number of intermediate outcomes, designed to assess the various pathways through which T4D may see impact on the outcomes under Research Questions 1 and 2. These outcomes link directly to the T4D intervention logic model (see Column C in Figure 3), and are associated with the social actions taken by intervention communities. Based on further analysis of qualitative data, the T4D team may amend this list of intermediate outcomes to include any additional pathways of impact that might emerge.

Logic Model Pathway	**Outcome definition**
Increased awareness, knowledge and improved community attitudes	Whether the respondent is able to correctly answer questions on healthcare practices**.**
Improved facility access (transportation, new facility, longer facility hours, outreach services)	**Build or request a new facility** - If (and how many) new facilities have been built/commissioned in the last 3 years.
Improved facility access (transportation, new facility, longer facility hours, outreach services)	**Community organized transportation** – • If the respondent reported using any form of transportation to go to the facility for delivery • If the respondent reported using an ambulance to go the facility for delivery
Improved facility access (transportation, new facility, longer facility hours, outreach services)	**Fix road** - Whether the respondent reported the fixing or building of new roads or bridges in the last 3 years, or the procurement of inflatable boats in the last 3 years, for traveling to the health facility.
Improved facility access (transportation, new facility, longer facility hours, outreach services)	**Travel time -** How long it took for the respondent to get to the facility
Improved facility access (transportation, new facility, longer facility hours, outreach services)	**Travel cost -** Amount of money spent by the respondent for transportation to get to the facility
Improved facility access (transportation, new facility, longer facility hours, outreach services)	**Mobile clinic or outreach services** - If (and how many) new mobile maternity clinics have been set-up in the last 3 years.

Improved facility access (transportation, new facility, longer facility hours, outreach services)	**Request ambulance** - Whether the facility has a functional ambulance.
Increased ability to pay	**Address cost of services** - Whether the respondent reported cost of care as a barrier to utilization of care, whether the respondent paid a fee for the delivery, total fees for the delivery reported by the respondent, whether the respondent (or spouse) has an insurance/health protection program enrolment
Increased ability to pay	**Raise community funds for delivery or other maternity costs, to support hospital patients (mothers)** - Whether the respondent was part of a woman's savings group to help save for costs associated with MNH care, whether the respondent reported using a community fund to help pay for costs associated with MNH care.
Bylaws, partnerships, or other interventions aimed at health system uptake	**By-laws -** If any of the sample villages in the facility's catchment area have a by-law or other local regulation, requiring/encouraging the uptake of MNH services, if the dispensary or health center referred to by the dispensary has a by-law or other local regulation, requiring/encouraging the uptake of MNH services.
Bylaws, partnerships, or other interventions aimed at health system uptake	**Midwife-TBA partnerships** - If there are midwife partnerships with TBAs.
Improved attitude, effort, trust of the provider (includes increased availability of provider)	**Monitor or complain about health facility staff performance** - The level of satisfaction reported by the respondent with the quality of MNH services, whether the respondent reported non-dignified care, provider neglect, or abuse, and provider perception of community feedback and information levels.
Improved attitude, effort, trust of the provider (includes increased availability of provider)	**Community/Village Health Worker residence in village** - If a CHW/VHW is i. Assigned to the village, and ii. Lives in the village
Improved facility cleanliness	Linked to actions around **cleaning the facility, or complaints about the cleanliness of the facility** - Level of cleanliness reported by the respondent, Observed level of cleanliness of the facility delivery room and toilet
Improved information transparency (cost, opening hours, etc.) or complaint mechanisms	**Cost transparency** - If cost information for delivery and other services is displayed publically

Improved information transparency (cost, opening hours, etc.) or complaint mechanisms	**Hours transparency** - If information on facility operating hours is displayed publically.
Improved information transparency (cost, opening hours, etc.) or complaint mechanisms	**Complaint mechanism** - i. If the facility has a complaint management system (e.g. a comment box) ii. If the facility conducts routine meetings with community members in the service area to identify how to improve quality of services
Improved provider knowledge	Linked to actions on **educating or training midwives**, this is defined as the number of questions the midwife can correctly answer questions on healthcare practices.
Improved facility infrastructure	**Fixing facility infrastructure** – Electricity, Telecommunications and computer/internet, Water, Delivery room (including privacy and beds), maternity home/resting place for women, toilet, communication (i.e. how would patients contact the facility), placenta pit.
Increased availability of drugs, supplies and other inputs	**Address shortage of medicines or supplies** - whether the facility has in stock essential medicines/equipment/vaccines for mothers and children, and basic supplies/equipment.
Increased or improved facility staffing	Number of staff in the MNH and birth unit, number of vacancies for staff.

The T4D team will also explore pathways linked to Research Question 4. In order for empowerment to transfer to the general community, the team hypothesizes that there must first be an increase in community activist empowerment, that the community activists must carry out social actions, and finally, that the general community must be aware that these actions took place (see Figure 5 under section 7.3).

The following intermediate outcomes measure community knowledge of the actions.

Logic Model Pathway	**Outcome definition**
Knowledge of social actions aimed at improving health outcomes	Number of social actions reported by respondents as having been carried out by Community Representatives
Knowledge of social actions aimed at improving health outcomes	Number of social actions that impacted the respondent positively

9 VARIABLE CONSTRUCTION

As mentioned above, outcomes pertaining to content of care will be combined into an index and treated as one outcome. Section 9.1 describes how that index will be constructed. The participation index will be constructed in the same way. The remainder of this section is relevant for all variables.

9.1 INDICES

The team will construct mean effect indices following the procedure outlined in Casey, Glennerster, and Miguel (2012)[56] which follows on Kling, Liebman, and Katz (2007)[57]. The steps involved in estimating the mean treatment effect are as follows:

1. Each outcome is first oriented so that higher values represent "better" values.
2. Then, each outcome is standardized by subtracting the mean of the outcome and dividing by the standard deviation of the control group.
3. Missing values are imputed at the treatment assignment group mean.
4. Finally, a summary index is compiled that gives equal weight to each individual outcome component. This index is then regressed as per the specifications described in Section 5.

The aforementioned approach weights each outcome component of the index equally. Anderson (2008)[58] weights each outcome component by the inverse of the appropriate element of the variance-covariance matrix (as measured in the control group), which "down-weights" outcome components that are highly correlated with each other. The team will check robustness using the weighted version and note any differences.

---

[56] Katherine Casey, Rachel Glennerster, and Edward Miguel, "Reshaping Institutions: Evidence on Aid Impacts Using a Preanalysis Plan," *The Quarterly Journal of Economics* 127, no. 4 (November 1, 2012): 1755–1812.

[57] Jeffrey R Kling, Jeffrey B Liebman, and Lawrence F Katz, "Experimental Analysis of Neighborhood Effects," *Econometrica* 75, no. 1 (January 1, 2007): 83–119.

[58] Michael L. Anderson, "Multiple Inference and Gender Differences in the Effects of Early Intervention: A Reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects," *Journal of the American Statistical Association* 103, no. 484 (December 1, 2008): 1481–95.

9.2 DON'T KNOW AND REFUSED

"Refused" will be coded as missing.

The treatment of "Don't know" is outcome-specific. In general, it will be coded as missing, but there are some exceptions. For instance, in the context of delivery, postpartum, and postnatal content of care, as well as the primary outcome on participation, "don't know" will be treated as a missing value, and imputed at the treatment assignment group mean. "Don't know" in response to knowledge questions will be coded as an incorrect answer. For birth preparedness, "don't know" will be coded as "no" (based on the assumption that if someone had done a specific preparatory activity, they would know of it). For all other variables, if more than 30% of the values are "don't know", the variable will be dropped from the analysis.

This will be followed generally, but we will keep the possibility open for imputing data in situations where the fraction of "don't know" responses is substantial, but not high enough to discard the variable entirely.

9.3 MISSING DATA FROM ITEM NON-RESPONSE

After recoding don't know and refused values, we will check for balance on missing values and test the sensitivity of our results to different assumptions on the missing data (due to non-response/DK, etc.). If necessary, due to various assumptions about missing data, we will create upper and lower bounds by recoding missing values for treatment as 0 and control as 1 and vice versa.

9.4 OUTLIERS

The majority of outcomes either are binary variables, indices/scales, or composed of a set of binary variables. Among primary and secondary outcomes, none is continuous. Among intermediate outcomes, only three variables are continuous[59].

For these variables, the team will first check that the reason for the outliers is not data entry error. If it is not, the analysis will be performed both including and excluding outliers, to

[59] The three variables are travel time, travel cost, and cost of services. This list may change slightly if the intermediate outcomes are amended. In any case, the procedure for dealing with outliers will be the same as described in Section 9.4.

check how sensitive the results are to the presence of outliers.

9.5 MULTIPLE BIRTHS

In cases of multiple births (e.g. twins), the last child to be born is measured. This is followed consistently for all variables.