



Transparency for Development: Evaluation Design Report

February 2015

Jean Arkedis*
Jessica Creighton**
Archon Fung**
Stephen Kosack†
Dan Levy**
Rohit Nainpally‡
Courtney Tolmie*

Acknowledgments: The William and Flora Hewlett Foundation, the U.K. Department for International Development, and the Bill and Melinda Gates Foundation provided generous funding for the work, which was originally conceptualized and brokered by the Transparency and Accountability Initiative. In particular, Kevin Bohrer, Tanweena Chowdhury, Peter Evans, Florencia Guerzovich, Brendan Halloran, Vanessa Herringshaw, Jenny Lah, Tom McEnroe, Nushina Mir, Alfonsina Penalosa, Helen Richards, and Tom Wingfield have been instrumental in shaping the project and providing substantive feedback along the way. The authors are very grateful to members of the T4D Technical Advisory Committee including Yamini Aiyar, Jessica Cohen, Jonathan Fox, Anuradha Joshi, Dan Posner, and Bill Savedoff for their comments and advice. Other colleagues have also generously given their time and constructive suggestions including Rikhil Bhavani, Jim McGuire, and Theodore Svoronos.

* Results for Development

** Harvard Kennedy School

† University of Washington

‡ Abdul Latif Jameel Poverty Action Lab

Forward

The Transparency for Development Evaluation Design Report describes the mixed methods evaluation design for the Transparency for Development project as of February 2015 – just before the start of the first data collection activities associated with the project (baseline data collection in Indonesia).

Since the writing of this report, a number of changes have been made to the design. For the most recent details, visit t4dproject.org.

Explanation of changes:

- Baseline data collection and design:
 - See the Transparency for Development [Baseline Report](#).
- RCT design, including declared primary outcomes and subgroups:
 - See the Transparency for Development [Pre-Analysis Plan](#) (V2).
- Key informant interviews:
 - In Tanzania only, the key informant interview (KII) process and sample was revised. Instead of conducting KIIs in 40 villages at the time of the 90-day follow up meeting, the research firm conducted KIIs in 24 villages at the time of the 90-day follow up meeting. The firm went back and collected a second round of KIIs in a subset of 10 of these villages at approximately 150 days.
- Endline data collection and design:
 - Sampling frame, endline household data collection: instead of interviewing 30 respondents per village, the survey firm interviewed an average of 30 respondents per village (6000 total, across the 200 villages in each country). The exact number of respondents per villages was determined by village population.
 - In Tanzania, the survey firm conducted the facility survey in 153 dispensaries, not 200.
 - In both countries, the survey firm conducted the facility survey in the sampled health facilities only; not additional health facilities that also serve the sample villages.
 - Endline community focus groups were administered to the community representatives and were administered in treatment villages only.

Contents

Executive Summary	5
I - Introduction	8
II – Description of the Interventions.....	10
I.1 Conceptual Framework and Logic Model of the Intervention.....	15
III – Key Research Questions.....	18
IV – Random Assignment Design.....	24
IV.1 - Unit of Random Assignment.....	24
IV.2 - Randomization Design and Implementation.....	25
IV.3 - Estimating Program Impacts	27
IV.4 - Sub-group Analysis.....	28
IV.5 - Statistical Power.....	29
V – Case Study Research	34
V.1 – Case Study Design Orientation.....	38
V.2 – Case Study Data Collection Layers and Integration with RCT Evaluation	39
VI - Sampling and Data Collection Plans	43
VI.1 - Sampling Frames.....	45
VI.2 - Data Collection Plan.....	46
VI.3 – Survey Instruments	50
VI.4 – Case Study Descriptive Elements	53
VII – Conclusion	63
References	65
Appendices.....	66

Figures

Figure 1- Intervention Activities	10
Figure 2 - Logic Model of the Intervention.....	16
Figure 3 - Community Actions and the Links to Intermediate Outcomes	17
Figure 4 - Random Assignment and Implementation	26
Figure 5 - Layers of Data Collection (the T4D Onion)	40

Tables

Table 1 - Information Collected.....	13
Table 2 - Primary and Secondary Intervention Outcomes.....	19
Table 3 - Intermediate Outcomes	22
Table 4 - Managing Threats to the Evaluation Design.....	27
Table 5 - Minimum Detectable Effects for Primary Outcomes (Overall Sample)	30
Table 6 - Minimum Detectable Effects for Primary Outcomes Sub-Group Analysis of Household Characteristics (at the Tercile Level).....	32
Table 7 - Minimum Detectable Effects for Primary Outcomes Sub-group Analysis of Cluster (Village) Characteristics (at the Tercile Level).....	33
Table 8 - Research Questions, Samples, and Methods.....	37
Table 9 - Data Collection Plan.....	43
Table 10 - Sample Survey Questions	51

Executive Summary

Generously funded by the William and Flora Hewlett Foundation, the U.K. Department for International Development, and the Bill and Melinda Gates Foundation and brokered under the Transparency and Accountability Initiative, the Transparency for Development (T4D) project will seek to understand the extent to which average citizens informed about their health care can identify, devise, and effectively implement solutions to the problems that prevent better public health service delivery and better health outcomes. This project draws from and builds upon the central ideas and practices from several related fields, including transparency and accountability, social accountability, and citizen participation, and is designed and executed by a team from the Harvard Kennedy School's Ash Center and the Results for Development Institute (R4D).

The goal of this report is to summarize progress made on two key activities of the T4D project:

I - Design and implement interventions with partner organizations in Indonesia and Tanzania aimed at improving maternal and neonatal health outcomes of households living in rural areas. The intervention is designed to involve seven activities:

1. Introductory Activities

Civil Society Organization (CSO)-employed facilitators will conduct necessary introductory meetings with village leadership, community health volunteers,⁴ and citizens.

2. Information Gathering

Household surveys will be used to collect information on the uptake, utilization, and contents of antenatal and labor and delivery services in intervention villages, as well as information on potential barriers preventing women in those villages from accessing these services. Facility surveys will assess availability of key inputs and staff.

3. Identification of Participants

A target of 15-16 intervention participants, referred to as “community representatives,” or CRs, will be recruited in each village to act as focal points in communities prior to, during, and after the intervention.

4. Community Meetings to Share Information and Develop an Action Plan

Two community meetings will be conducted with the community representatives. On the first day, communities will discuss the contents of a “community scorecard” reflecting the findings from the household and facility surveys. On the second day of meetings, communities will be asked to come up with a plan of

⁴ Indonesia only

actions they intend to take to improve maternal and neonatal health – the “social action plan.”

5. **Open Public Meeting**

After the CRs have developed the social action plan, the facilitator and community representatives will present the plan back to the larger community in the form of an open public meeting.

6. **Social Actions**

The CRs will begin carrying out the activities detailed in the social action plan. These activities are intended to be carried out independently by the CRs, without the help of the CSO facilitator.

7. **Follow-up Meetings**

A series of three follow up meetings over a period of 90 days will be conducted to check-in with the community to see how they are progressing with the social actions and help revise their approach if the original approach is not working.

II - Conduct an evaluation to assess whether the interventions improve health outcomes and under what conditions. The key research questions the evaluation will seek to answer are the following:

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 citizens’ perceptions of **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?

The evaluation will use a mixed methods approach to answer these questions. Two randomized control trials (RCTs), one each in Tanzania and Indonesia, will be undertaken to evaluate the intervention’s effects on health care utilization, content and outcomes, and on community empowerment. These RCTs will be used primarily to answer research questions 1-4.

In addition, extensive case studies of a subset of the treatment and control communities will allow a much richer understanding of the answers to questions 1-4 and to answer questions 5 and 6. In these case study communities, focus groups, informant interviews,

systematic coding of meetings, empowerment surveys, facilitator reports, and ethnographic methods will allow an understanding of the context in which the interventions occurred and enable us to process-trace exactly how the interventions triggered—or failed to trigger—improvements in health care and changes in power dynamics and community relations.

I - Introduction

Recent decades have seen vast increases in health expenditures and expansion of services across much of the developing world, but improvements in health outcomes have been uneven. Over the last ten to fifteen years, child survival has improved dramatically in both Tanzania and Indonesia, while death within the first month of life is largely unchanged and maternal mortality is high. Evidence suggests that this is due at least in part to problems with the delivery and uptake of maternal and neonatal health (MNH) services, which are predominantly delivered by the public sector. Problems exist at a number of levels: within the community (lack of information, awareness, and cultural practices), at the facility (lack of effort of providers, poor management, waste or leakage of drugs and equipment), and within the larger health system (supply chain problems, human resource constraints, poor infrastructure, access issues).

Citizen-led action—directed toward others in the community, providers and facilities, or other political or health system actors—may be able to address some of these problems. However, the existing evidence base evaluating interventions that seek to facilitate or encourage citizen-led action is mixed. This literature also largely leaves untested the idea that context and community-specific “fit” are key to successful community action. No existing studies rigorously evaluate interventions that allow communities to choose the sort of actions they will undertake: whether, for example, to work *with* providers, to *put pressure* on providers, political actors, or other policy makers, or to undertake self-help actions.

Generously funded by the William and Flora Hewlett Foundation, the U.K. Department for International Development, and the Bill and Melinda Gates Foundation and brokered under the Transparency and Accountability Initiative, the Transparency for Development (T4D) project will seek to understand the extent to which average citizens informed about health problems can identify, devise, and effectively implement solutions to the problems that prevent better public health service delivery and better health outcomes. This project draws from and builds upon the central ideas and practices from several related fields, including transparency and accountability, social accountability, and citizen participation, as well as a range of academic disciplines, and is designed and executed by a team from the Harvard Kennedy School’s Ash Center and the Results for Development Institute (R4D).

The overriding goal of the Transparency for Development project is to generate actionable evidence for practitioners, researchers, and other stakeholders working to improve health, accountability, and citizen participation.

For the purposes of this evaluation report, we focus on the following two key activities of the project:

I – The design of interventions in Indonesia and Tanzania aimed at improving health outcomes of women and children living in rural areas. The interventions will provide

information to citizens about health in their community and will seek to mobilize them to design and implement a social action plan aimed at improving health outcomes for the community.

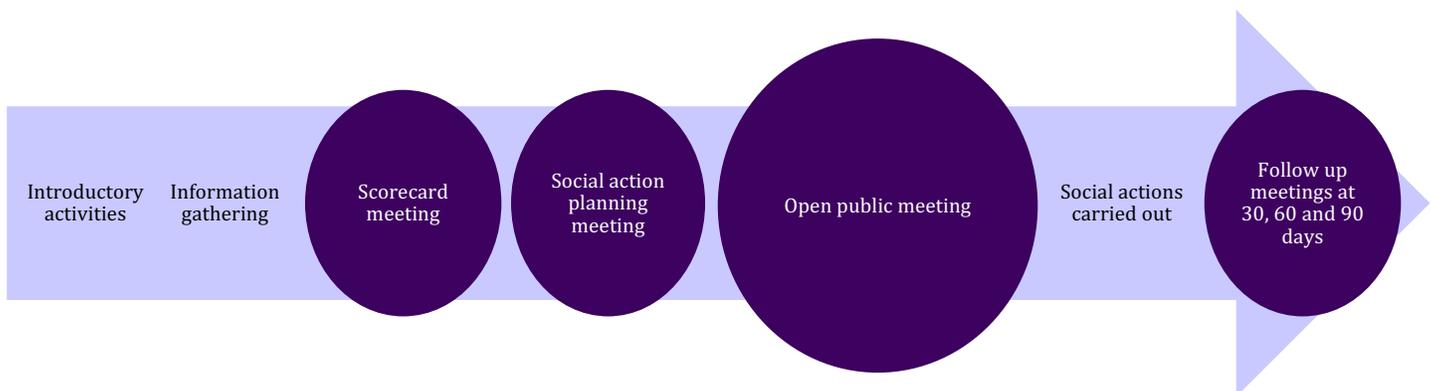
II – A mixed-methods evaluation to assess whether the interventions improve health outcomes and citizen empowerment and, if so, how and under what conditions. The evaluation will rely on a mix of quantitative and qualitative methods to estimate the impacts of the intervention, assess its implementation, and understand the pathways through which the intervention might have affected the outcomes of interest.

This report is organized as follows. Chapter II provides a description of the interventions, including a logic model that suggests possible pathways for the interventions to affect outcomes. Chapter III describes the key research questions and specifies the principal outcomes that the impact evaluation will seek to measure. Chapter IV describes the randomized controlled trial (RCT) design. Chapter V describes the elements of the qualitative case-study research design, and Chapter VI presents the sampling and data collection plan for both the quantitative and qualitative parts of the evaluation. Chapter VII concludes.

II – Description of the Interventions

The T4D intervention aims to improve village-level MNH in rural communities using a modified version of a “community scorecard.” The intervention comprises seven total activities; six undertaken by the Civil Society Organization (CSO) partners: (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 an open public meeting; and (7) a series of facilitated follow-up meetings.. In addition to these activities led by our CSO partners, community representatives will (6) independently undertake social actions to address health problems that they identify as priorities during the meetings led by CSO facilitators. These intervention components are described below and illustrated in Figure 1. Since the interventions will be very similar for Indonesia and Tanzania, this chapter describes the two interventions as a single intervention while noting areas in which the two interventions differ. Minor changes may still be made to the intervention based on the final results of pilots and intervention redesign discussions with our partners.

Figure 1- Intervention Activities



1. Introductory activities

At the start of the intervention, CSO-employed facilitators enter assigned villages and meet with village leadership, community health volunteers,⁵ and citizens. At these meetings, the facilitators explain the intervention and its aims, identify survey respondents, and identify potential intervention participants, or “community representatives.” This introduction is also intended to encourage ownership of the project by community members.

⁵ Indonesia only

2. Information gathering

Next, the facilitator collects data on maternal and neonatal health as well as health service delivery. The facilitator conducts a beneficiary survey with 20-30 women who have given birth in the last two years, to assess health outcomes of mothers and babies, utilization and coverage rates of services, and health facility performance as experienced by the respondent. The facilitator also conducts a simple facility survey to assess infrastructure, cleanliness, and availability of drugs and supplies.

Table 1 below summarizes the information that is collected. As Table 1 shows, information is collected on antenatal care (ANC), delivery services, and postnatal and postpartum care. The second column in the table (“health levers”) includes statistics to quantify the proportion of women receiving ANC and delivering in a facility with a skilled birth attendant, and the proportion of women and babies receiving postnatal care. The remaining columns include information on potential barriers (“Barriers”) preventing women from accessing these services, such as knowledge and cultural barriers, cost and other access barriers, and barriers associated with the facility itself and the provider’s actions. The information collected will differ slightly across countries due to differences in the major maternal and neonatal health problems and barriers in the two countries.

3. Identification of intervention participants

In parallel, the facilitator identifies 15-16 local community members with a particular interest in making improvements in MNH. These individuals are expected to participate in the community scorecard and social action meetings (described in the next section), as well as carry out the social actions. Known as “community representatives” or CRs, these participants are recruited based on a number of characteristics, including: leadership potential, time and willingness to volunteer, and/or enthusiasm about improving the village. Formal leaders and health workers are excluded.

4. Facilitation of meetings to share information and develop a social action plan

Community representative meetings take place over two days and are organized around two sets of activities: sharing the collected information in the form of a “community score card” and developing a social action plan to address some of the major problems revealed by the information.

Day 1: Scorecard Meeting

The survey information on health levers (see Table 1) is aggregated and presented to the community representatives along with motivating information describing the high rates of maternal and neonatal death and the link between the health levers and improving maternal and neonatal health. Facilitators also use the information collected on barriers (see Table 1) to help highlight potential reasons why communities may not be realizing higher rates of ANC, delivery, and postnatal care.

Community representatives are presented with short vignettes of actions that other communities have taken to improve service delivery, uptake, or both (“social action stories”). These social action stories have two purposes. First, they are intended to start the community thinking that it is possible for the community themselves to undertake individual and social actions to improve health care, since many other communities have done just that. Second, they introduce a variety of different ways for communities to try to improve uptake and care, in an effort to start community members thinking about which might be appropriate to their context. The social action stories highlight actions directed toward members of the community (self-help), providers and the facility, as well as other policy makers and politicians. The stories are intended to stimulate communities to consider both collaborative (e.g. working with providers to improve services) and more oppositional (e.g. complaint or supplication) approaches.

Day 2: Social Action Planning Meeting

On day two, community representatives are led through a process of developing a plan of actions they intend to take to improve maternal and neonatal health – the social action plan. The social action stories are again discussed, but the facilitator emphasizes that communities need not rely on these; they have complete flexibility to choose the types and targets of their actions. Community representatives are asked to formulate at least some actions that can be implemented and lead to improvements in the short term (90 days), as well as to devise longer-term actions. They are asked to come up with a number of specific activities and to assign a specific person as responsible for leading each action.

5. Sharing the action plan with the greater community during an open public meeting

After the social action plan is developed, an open public meeting is held to share an abbreviated version of the community scorecard and the social action plan, to invite comment and additions, and to provide opportunities for other community members to volunteer to participate in decided-upon actions.

6. Community-led social action

After these initial meetings, the community representatives are expected to independently carry out the social actions. The CRs are not given monetary or other resources for the actions.

7. Facilitated follow up meetings

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

Table 1 - Information Collected

Continuum of Care	Health Levers	Barriers ⁶		
		Knowledge, Awareness, and Cultural	Cost/Access	Facility/Provider
ANTENATAL CARE (ANC) ⁷	<p>Proportion of pregnant women who initiate ANC in the first 12 weeks of pregnancy</p> <p>Proportion of pregnant women receiving 4 or more ANC visits</p> <p>Proportion of pregnant women creating a birth preparedness plan</p>	<p>Proportion of women who know they should receive ANC care within the first 12 weeks of pregnancy</p> <p>Proportion of women who report that they did not think it was important for them to receive ANC</p> <p>Proportion of women who report a lack of support from family/others as a barrier preventing them from receiving ANC</p> <p>Proportion of women who report lack of male support/permission as a barrier preventing them from receiving ANC</p> <p>Proportion of women who reported superstition/fear of witchcraft as a barrier preventing them from seeking ANC</p>	<p>Proportion of women who report paying anything for their ANC</p> <p>Proportion of women who report cost as a reason why they did not attend ANC</p> <p>Proportion of women who say that lack of transport is a barrier preventing them from seeking ANC</p> <p>Proportion of women who say that distance to the facility is a barrier preventing them from seeking ANC</p>	<p>Negative attitude of healthcare provider</p> <p>Perceived absenteeism or availability of staff at the facility</p> <p>Cleanliness of the facility</p> <p>Presence of a separate/private delivery room</p> <p>Toilet at the facility</p> <p>Stock out of key medicines or supplies</p> <p>Waiting time to see midwife/health facility staff</p> <p>Poor midwife knowledge or effort</p> <p>Placenta pit</p> <p>Availability of a female midwife</p>
DELIVERY	<p>Proportion of women delivering in the health facility</p> <p>Proportion of women delivering with a skilled attendant</p>	<p>Proportion of women who agree with the statement that it is safer to give birth in a facility</p> <p>Proportion of women who report a lack of support from family/others as a barrier preventing them from delivering in a facility</p> <p>Proportion of women who report lack of male support/permission as a barrier preventing them from delivering at a facility</p>	<p>Percentage who report paying a bribe for labor and delivery services</p> <p>Average out-of-pocket cost per facility birth</p> <p>Proportion of women who say that lack of transport is an important barrier preventing them from giving birth in a facility</p>	<p>Broken/missing equipment</p> <p>Refrigeration/electricity at facility</p> <p>Water availability</p> <p>Facility operational hours not observed or unclear</p> <p>Information on cost not observed or unclear</p> <p>No female healthcare worker</p>

⁶ The specific barriers information is still being finalized and will vary between Indonesia and Tanzania.

⁷ We are considering focusing on delivery and postnatal care in Indonesia, in which case we would not collect information on ANC.

		<p>Proportion of women who feared being operated on (Caesarean section) if they delivered in a facility</p> <p>Proportion of women who reported superstition/fear of witchcraft as a barrier preventing them from delivering in a facility</p> <p>Proportion of women who say that they would prefer a TBA, even if money was not a factor</p> <p>Proportion of women who made a birth preparedness plan</p>		
<p>POSTNATAL AND POSTPARTUM CARE</p>	<p>Proportion of babies who get a check-up from a health worker in the first week of life</p> <p>Proportion of women who get a check-up from a health worker within one week of giving birth</p>	<p>Proportion of women who know that babies should have a check-up within the first week</p> <p>Proportion of women who report that they do not think postnatal care was important</p> <p>Proportion of women who report a lack of support from family/others as a barrier preventing them from receiving postnatal care</p> <p>Proportion of women who report cultural reasons for not taking the baby out of the house for a certain period as a reason why they did not seek postnatal care</p>	<p>Proportion of women who say that distance to the facility is an important barrier preventing them from seeking postnatal care</p> <p>Proportion of women who report paying anything for their postnatal care</p> <p>Proportion of women who report cost as a barrier preventing them from receiving postnatal care</p> <p>Proportion of women who say that lack of transport is a barrier preventing them from bringing their babies for a postnatal check-up within the first week</p>	

I.1 Conceptual Framework and Logic Model of the Intervention

Figure 2 illustrates how the intervention is hypothesized to affect health outcomes. To have an impact, at a minimum the community must understand and be motivated by the information, develop a plausible social action plan, and successfully carry it out. There are three main pathways through which this process may improve health outcomes (E. Health Outcomes in Figure 2): 1) the proportion of people receiving services increases (increased utilization); 2) the quality of services delivered through existing channels improves (improved content of care); and/or 3) people who were receiving lower quality care at one outlet choose to seek care at a higher quality outlet.

This intervention has primarily been designed to trigger the first two of these—collective action targeted at improving either service utilization (D1 in Figure 2), the clinical content of services (D2 in Figure 2), or both. Though there is nothing preventing communities from seeking care at different outlets, the information component of the intervention does not help them to assess the relative quality of different health facilities.

In the intervention, communities are asked to devise a social action plan and carry out community actions (B in Figure 2). Figure 3 expands upon the range of options communities might pursue. It provides more detail on the types of social action the community might take, the targets of those actions, and their link to intermediate outcomes. Collective actions can either target community members (self-help actions), the provider or staff at a facility, local politicians or policymakers, others in the service delivery chain, other political leaders or policymakers (such as those at the regional or national level). For example, communities can directly try to improve utilization by implementing information campaigns (self-help actions to improve awareness, knowledge, or attitudes related to maternal and neonatal health) or undertake other types of self-help initiatives such as organizing transportation pools for pregnant women so they can deliver in a distant facility (self-help actions to increase access). Types of action 2, 3, 4, and 5 all describe strategies that communities could direct toward front-line service providers to improve patient's experience in the health facility and her interaction with the provider in some way (Intermediate Outcomes C3-6). For example, this could involve strategies to eliminate illegal fees and reduce the cost of services, incentivize or compel service providers to exert more effort, work more hours, and/or improve their attitude toward patients, etc. The community might also work with providers to improve the management of the facility in some way that reduces waiting time and/or some other inefficiency in patient care. Alternatively, communities could pursue a “longer route” option (Types of Action 5 and 6) by working with or putting pressure on district or other health officials who supervise front-line service providers to improve the availability of drugs and equipment, the quality or number of staff available, or other key inputs (C7-8) or, alternatively to hold front-line service providers accountable for improvements in the patient's experience within the health facility (C3-6).

Figure 2 - Logic Model of the Intervention

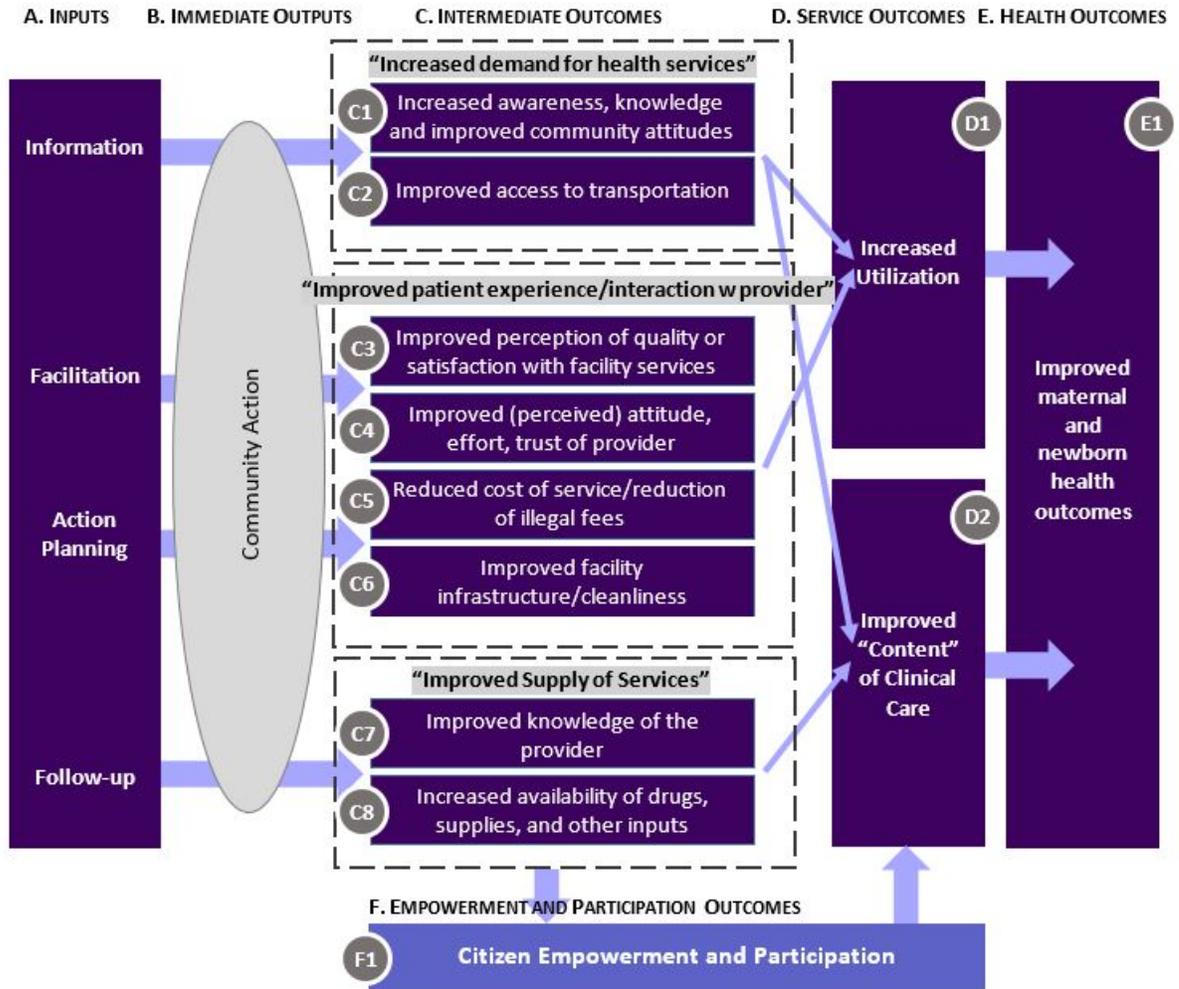
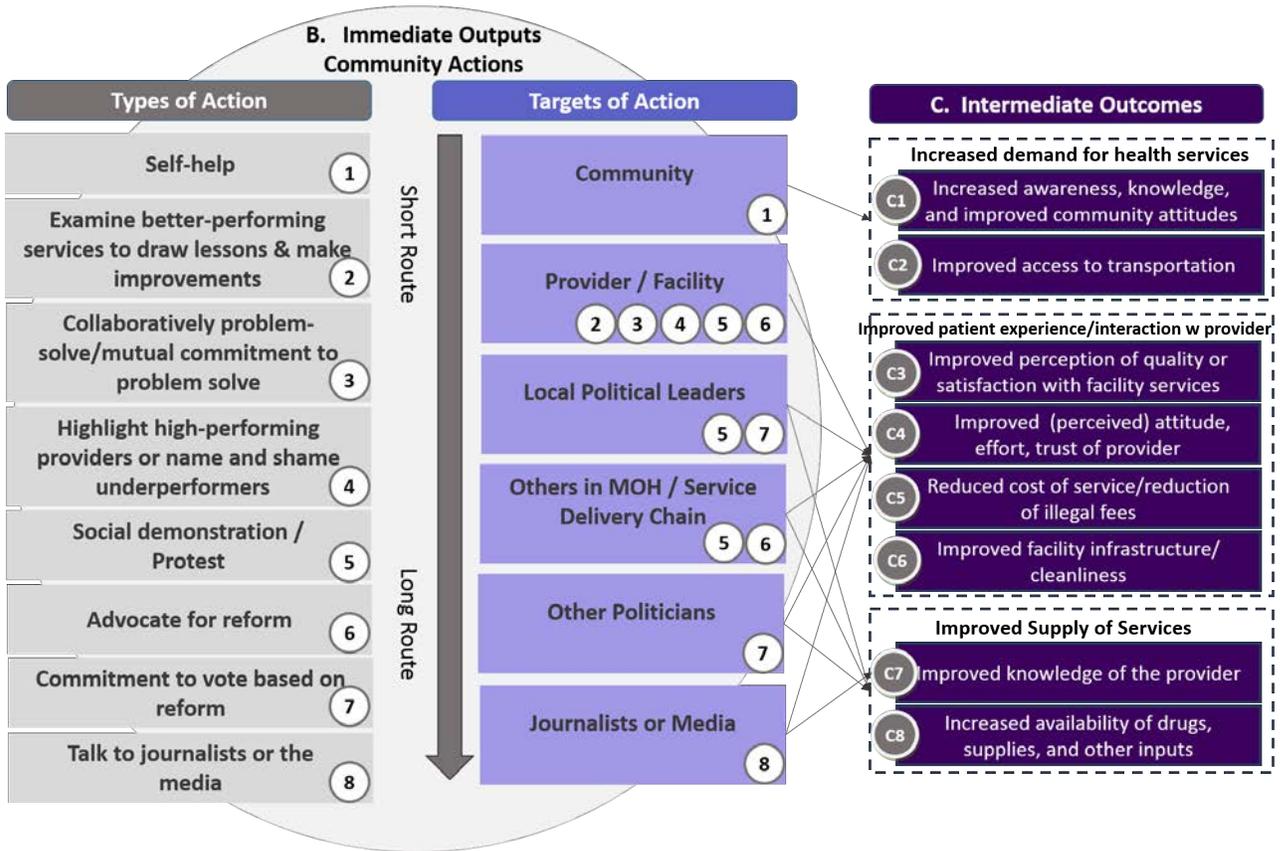


Figure 3 - Community Actions and the Links to Intermediate Outcomes



III – Key Research Questions

The key research questions the evaluation will seek to answer are the following:

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 **clinical 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 citizens' perceptions of **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?

The evaluation will use a mixed methods approach to answer these questions. Two randomized control trials (RCTs), one each in Tanzania and Indonesia, will be used primarily to answer research questions 1-4. In-depth case studies of a subset of the treatment and control communities will be used to provide a much richer understanding of the answers to questions 1-4 and to answer questions 5 and 6. These case studies will combine focus groups, informant interviews, systematic coding of meetings, empowerment surveys, facilitator reports, and ethnographic methods, all to allow an understanding of context and to process-trace exactly how the interventions triggered—or failed to trigger—improvements in health care and changes in power dynamics and community relations.

Table 2 lists the primary outcomes on which the RCT portion of our evaluation will seek to estimate impacts for research questions 1-4.

Table 2 - Primary and Secondary Intervention Outcomes

Research Question	Corresponding Column/Box in Figure 2 – Logic Model	Primary Outcomes ⁸	Secondary Outcomes
1. What is the effect of the intervention on the utilization of health care services related to maternal and child health?	D1. Service Outcomes: Increased Utilization	<ul style="list-style-type: none"> • Proportion of women delivering at a birth facility • Proportion of women delivering with a skilled attendant • Proportion of women receiving 4 or more ANC visits • Proportion of women attending first ANC visit within the first 3 months of pregnancy • Proportion of women receiving postpartum check within 7 days following delivery • Proportion of babies receiving postnatal check within 7 days following birth 	
2. What is the effect of the intervention on the content of health care services related to maternal and child health?	D2. Service Outcomes: Improved Clinical Content of Care		ANC content/quality index ⁹ including the proportion of women receiving the following interventions during ANC: <ul style="list-style-type: none"> ○ Iron tablets or syrup ○ Anti-malarial drugs¹⁰ ○ Took deworming medication¹¹ ○ Informed of signs of pregnancy complications ○ Blood pressure taken ○ Urine sample taken

⁸ In Indonesia, we are considering focusing exclusively on delivery and postnatal/postpartum care, in which case we would not look at the following primary outcomes associated with ANC:

- Proportion of women receiving 4 or more ANC visits
- Median months pregnant at first visit (for those with ANC)
- ANC content/quality index

⁹ We will create the indices using one of the standard methods used in the literature (e.g. mean effects index (Kling, Liebman, and Katz 2007), principal component index (Casey, Glennerster, and Miguel 2012), etc.). The final selection of which index we report as our principal index of interest will depend on the nature of the data that we collect; however, we will check for and report on the robustness of our findings to the use of alternative indices.

¹⁰ Tanzania only

¹¹ Ibid.

			<p>Maternal postpartum content of care index including the proportion of women who received the following interventions during a postpartum visit:</p> <ul style="list-style-type: none"> ○ Examined body ○ Checked breasts ○ Counseled on danger signs for newborns ○ Counseled on danger signs for mothers ○ Counseled on breastfeeding ○ Counseled on family planning <p>Baby postnatal care content of care index including the proportion of babies receiving the following interventions during first postnatal visit:</p> <ul style="list-style-type: none"> ○ Examined body ○ Weighed baby ○ Checked cord ○ Checked for danger signs ○ Immunizations
3. What is the effect of the intervention on health outcomes ?	E. Health Outcomes (child)	<ul style="list-style-type: none"> ● Age-for-weight (z) scores ● Height-for-age (z) scores 	<ul style="list-style-type: none"> ○ Infant mortality¹² ○ Birth weight
	E. Health Outcomes (mother)		<ul style="list-style-type: none"> ○ Mental health of mothers¹³
4. What is the effect of the intervention on citizen empowerment and efficacy , both perceived and actual?	Citizen Empowerment and Participation	<ul style="list-style-type: none"> ● Perceptions of power to improve life and village ● Perceptions of responsiveness of health facilities and health providers to community needs ● Perceptions of responsiveness of state officials to community needs ● Participation in communal collective action 	

¹² Assuming that the intervention impacts the proportion of women that give birth at a facility, it is conceivable that infant mortality rates will be reduced. Given this is a rare outcome even in the poor villages where the intervention will take place, it is unlikely that the study will have sufficient statistical power to detect effects on this outcome. But given this is a policy relevant outcome, we will try to assess impact.

¹³ If the intervention is successful at improving mothers' control over the quality and safety of health care during birth, it may also improve mothers' mental health, in particular reducing their vulnerability to depression. Perceptions of control are strongly related to depression, and the period around pregnancy is one of the periods when women are most likely to have a major depressive episode. Depression is the 10th leading cause of disability in the world (higher than Malaria, TB, anemia, or diabetes) and it is a recurrent illness: once a person experiences a major depressive episode they are far more likely to have one again. Depression in mothers is also strongly related to poor physical health of children: higher risk of growth retardation and diarrheal diseases, as well as poor developmental outcomes including cognitive impairment.

Note that while it is possible for the intervention to have impacts on a wider set of outcomes, we are purposefully restricting our attention to a small set to limit the chances of falsely declaring impacts that result from multiple testing.¹⁴ We will examine impacts on additional outcomes, but we will declare *a priori* the outcomes that will be markers of whether the intervention succeeded in terms of having impacts on the intended beneficiaries. The final set of primary outcomes will be clearly specified in a pre-analysis plan that we will register on the American Economic Association's RCT [Registry](#).

Finally, the last column in Table 2 lists a set of secondary outcomes that our intervention may affect. For these outcomes, the links to our intervention are not as strong as with the primary outcomes of interest, and hence it is possible that the evaluation design will not have sufficient statistical power to detect impacts on these outcomes. Thus these are not outcomes that we will *a priori* declare to be markers of whether the intervention succeeded.

In addition to measuring impacts of the interventions on the primary outcomes above, we will also estimate impacts on several intermediate outcomes that are derived from the logic model (C1-C8 in Figure 2 and Figure 3). These are listed in Table 3, below. These impacts, combined with the case-study research, will shed light on the mechanisms that could be responsible for the existence (or lack) of impacts on the main outcomes of interest.

¹⁴ As the number of hypotheses tested in a given study increases, the probability of a Type I error (i.e. incorrect rejection of the null hypothesis of no impact, in this case) increases. For the substantive implications of multiple hypothesis testing, see Anderson (2008). When testing multiple hypotheses, we will employ statistical corrections, such as the Familywise Error Rate correction and False Positive Discovery rate controls, to guard against Type I errors. However, these corrections tend to reduce the power of the study insofar as they impose more stringent requirements for declaring statistical significance. This leads to a greater risk of Type II errors (i.e. incorrect acceptance of the null hypothesis of no impact). In sum, we chose a limited set of key outcomes in order to be able to balance the likelihoods of committing two types of statistical errors that trade off against each other.

Table 3 - Intermediate Outcomes¹⁵

Logic Model Category		Intermediate Outcome ¹⁶
Increased Demand for Health Services	C1. Increased household knowledge, awareness, and attitudes related to MNH care and services	<p>Knowledge index of women regarding:</p> <ul style="list-style-type: none"> • Timing and frequency of ANC care • Warning signs during pregnancy • Birth preparedness • Timing and reason for postnatal care for babies <p>Attitude index of women regarding:</p> <ul style="list-style-type: none"> • Whether ANC care is necessary if a patient had no complications in a previous pregnancy • Whether it is safe to wait at home until complications arise before going to the health facility for delivery • Relative safety of giving birth at home with a TBA to giving birth in a health facility.
	C2. Improved access to transportation	<ul style="list-style-type: none"> • Proportion of women reporting that transportation as a barrier for delaying/not seeking otherwise desired services • Median amount of time (in minutes) it takes to travel to the health facility • Median cost of transportation to the health facility
Improved patient experience/interaction with provider	C3. Improved perception of quality or satisfaction with facility services	<ul style="list-style-type: none"> • Patient rating of delivery quality • Patient rating of quality of most recent visit to the target facility
	C4. Improved (perceived) attitude, effort, and trust of provider	<ul style="list-style-type: none"> • Patient rating of provider communication during delivery • Patient rating of degree of respect shown by provider during delivery • Patient rating of provider communication during most recent visit • Patient rating of degree of respect shown by provider during most recent visit • Patient overall rating of their degree of trust in nurses, midwives, or other staff at local facility
	C5. Reduced cost of service/reduction of illegal fees	<ul style="list-style-type: none"> • Proportion of women reporting that cost is a reason for not seeking ANC/delivery/post-partum care • Proportion of ANC/delivery/post-partum care visits where women paid a fee during the most recent pregnancy • Proportion of households reporting payment at their most recent visit to the local facility
	C6. Improved facility management/cleanliness	<ul style="list-style-type: none"> • Patient perceptions of waiting times • Patient perceptions of facility efficiency • Observed cleanliness • Patient perceptions of cleanliness • Absenteeism

¹⁵ Note that the information on this table is still being finalized.

¹⁶ We are considering focusing exclusively on labor and delivery and post-partum care in Indonesia, in which case we would not look at intermediate outcomes related to ANC.

Improved supply of services	C7. Improved knowledge/attitudes of the provider	<ul style="list-style-type: none"> • Knowledge index of providers regarding: • Timing, frequency, and clinical contents of ANC care • Warning signs during pregnancy • Timing and contents of post-natal care for mothers and babies • Appropriate post-natal vaccinations • Attitudes of provider regarding: • When ANC care is necessary if a patient had no complication in a previous pregnancy • Whether it is safe to wait at home until complication arise before going to the health facility for delivery
	C8. Increased availability of drugs, supplies, and other inputs	<ul style="list-style-type: none"> • Essential MNH drug stock outs (last 3 months) • Essential MNH supply stock outs (last 3 months) • Patient perception of drug availability

As indicated above, because the social action plan in the intervention is open, communities might choose different pathways to pursue improvements in health outcomes. Depending on the extent that they do, it is possible that impacts on some of these intermediate outcomes will not be detected, not because they did not exist but rather because they correspond to pathways that were not activated by a large enough number of communities. This is clearly a tradeoff between choosing an intervention design feature we believe is a key ingredient of a successful and sustainable intervention (the open social action) and choosing an evaluation design that has the highest chance of pinning down quantitatively the mechanisms behind the impacts.¹⁷ Initial research and conceptual work to prepare for this project led us to deliberately choose the former, though we are aware that different researchers could have made a different choice.

¹⁷ For example, randomly assigning kind of social actions to different communities would have allowed us to quantitatively evaluate these mechanisms more precisely, but would have compromised the open social action approach.

IV – Random Assignment Design

The impacts of the interventions will be estimated using Randomized Controlled Trials (RCTs). 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. This section describes the design of the RCTs.

IV.1 - Unit of Random Assignment

The evaluation will focus on a selected sample of 200 villages from each of two countries: Indonesia and Tanzania. In Indonesia, the sample will be drawn from two provinces (Banten and South Sulawesi); in Tanzania, the sample will be drawn from two regions (Tanga and Dodoma). The unit of randomization will be at the health facility level.

In Indonesia, we define a health facility as a puskesmas. A puskesmas is a public health center that operates at the subdistrict level (in larger subdistricts there may be more than one puskesmas). Puskesmas are the lowest level public health service center overseen by the Indonesian government. They provide comprehensive basic health services, such as health promotion, prevention, treatment and rehabilitation. Patients with more serious diseases are transferred to a local hospital. Each puskesmas serves up to 30,000 people from 5-10 villages. A puskesmas may be supported by a network of additional health centers, such as puskesmas pembantu (assistant puskesmas) and puskesmas keliling (mobile puskesmas), that are directly overseen by the puskesmas.

In Tanzania, we define the health facility as a dispensary. In Tanzania's pyramidal health structure, a dispensary is the lowest level health facility, serving a catchment area of approximately 6,000 to 10,000 people across 2-8 villages. Dispensaries should be staffed with a medical assistant, a public health nurse or nurse midwife, and health assistant to provide both preventative and curative outpatient services. Complicated patient cases are referred to nearby health centers or hospitals. Some of the services provided by dispensaries include: health education, maternal child health services (antenatal and postnatal care), delivery services for uncomplicated pregnancies, treatment and immunization services to children, as well as outreach to villages for the provision of health services.

For both puskesmas and dispensaries, catchment areas will be defined as all villages primarily served by each facility. To determine the catchment area, we will obtain the official list of villages served by each facility and verify these lists with each facility to ensure accuracy.

IV.2 - Randomization Design and Implementation

We will conduct a two-arm study in each country. We will identify a list of 200 eligible puskesmas in Indonesia and 200 eligible dispensaries in Tanzania. One village will be randomly selected from the catchment area of each health facility to be surveyed two times – once prior to the intervention, and once afterwards.¹⁸ Half of the surveyed villages will be randomly selected to receive the intervention treatment. The other half will serve as control villages.

The eligibility criteria for the health facilities (puskesmas or dispensary) is the following:

- They must be located in either Banten¹⁹ province or South Sulawesi province, Indonesia, or Tanga or Dodoma region, Tanzania.
- They must be located in rural areas.
- They must not be currently, or have recently been, involved in a similar transparency and accountability demand-side intervention.

The eligibility criteria for the villages is the following:

- They should be located within 7 km of the health facility.²⁰
- More than 2,000 people should live in the village²¹

To implement random assignment, we will do the following (see also Figure 4):

1. Compile list of health facilities in the regions/provinces that are part of this study
2. Apply eligibility criteria to the list of health facilities
3. Choose 200 eligible centers at random
4. For each health center, determine the villages that are in its catchment area. To do so, we will employ one or more of the following methods:
 - a. Find this information from official data sources
 - b. Find this information using GPS locations of village centers and health facilities
 - c. Call each of the centers and ask them for a list of villages that they serve
 - d. Collect these data during the baseline
5. Among the list of villages served by the facility, apply the eligibility criteria described above

¹⁸ See “Data Collection Plan: Timeline” for more details on the timing of the two survey rounds.

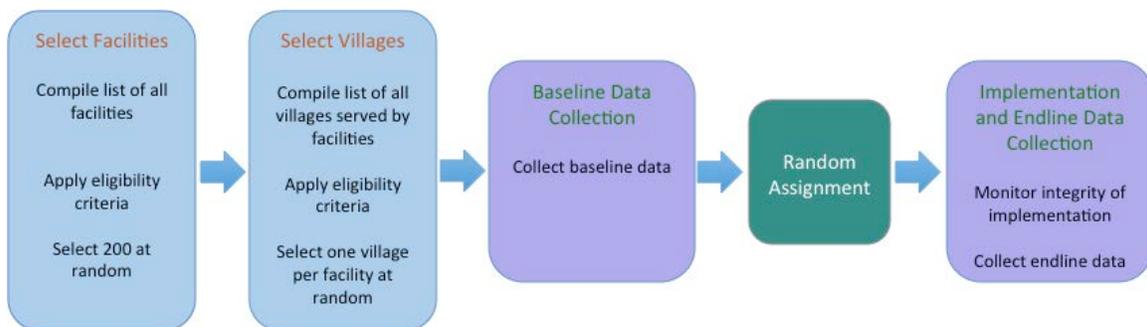
¹⁹ Excluding the Baduy area because this area is historically closed to CSOs and other perceived outsiders.

²⁰ The goal with this criteria is to eliminate the furthest 25% or so of villages, as this intervention is unlikely to be very effective in the remotest villages. We will use data obtained during the baseline facility survey and village listing exercise to determine the precise distance from the facility to use as a maximum radius.

²¹ This is to ensure that the sample size of households eligible for the study is large enough given our statistical power requirements. We will use the data obtained during the baseline survey to determine the precise population criteria.

6. Among those villages eligible for the intervention, select one at random from each catchment area²²
7. Collect baseline data for these 200 villages
8. Randomly assign these 200 villages to treatment and control groups. At the time of this writing , we have not decided on whether we will stratify the random assignment. If we do, we will collect data on the stratifiers at baseline.
9. Verify that assignment looks random by conducting baseline equivalence tests
10. Implement the program in treatment villages. We will employ measures to make sure that fidelity to treatment is respected and to manage threats to the integrity of our evaluation design (see Table 4 below)
11. Collect follow-up data for treatment and control villages

Figure 4 - Random Assignment and Implementation



As with any evaluation design, this one implies risks. Table 4 describes the main risks we believe this evaluation design entails and our plans for mitigating these risks.

²² Recognizing that some villages might appear in more than one catchment area, we will apply an algorithm that will seek to avoid a scenario in which our final sample ends up containing a treatment village and a control village in the catchment area of the same facilities. We will do so by eliminating certain facilities from our sample rather than certain villages, given that we want to keep in our sample villages that are served by more than one facility.

Table 4 - Managing Threats to the Evaluation Design

Risk	Description	Mitigation Strategy
Insufficient statistical power	Intervention may have an effect but evaluation may not detect it statistically	We have performed conservative power calculations at the outset to ensure that evaluation can detect policy-relevant effects (please see section IV.5).
Spillovers	The treatment spills over in control units, threatening the ability of the evaluation to credibly estimate the impact of the intervention	We plan to mitigate spillover effects by randomizing at the level of the facility catchment area. We will select communities with sufficient geographic distance from each other so that there is limited opportunity for the population in the catchment area of a control facility to be exposed to treatments occurring in another community (see section IV.2).
Data collection risks	Data collection quality is compromised by poor performance of the data collection firm	We have selected our survey firms using a competitive request-for-proposal and interview process and believe we have selected two high quality data collection firms at a competitive price. T4D study coordinators based in Tanzania and Indonesia will closely monitor the firms during questionnaire design, survey piloting, interviewer training, actual data collection, data entry, data cleaning and data delivery.
Implementer and other stakeholders are not faithful to the random assignment	Due to political pressure, misunderstanding, or other reasons, the implementer implements the intervention in control areas or fails to implement in treatment areas	The evaluation design and the need for fidelity to the random assignment was emphasized during the implementation partner selection process, and willingness to participate in a randomized evaluation was a selection criteria. In advance of program scale up a presentation to all staff will highlight the rationale for this approach.
Implementer does not fully implement the intervention in the full sample of treatment communities	Implementer has problems in delivering the intervention to all communities in our treatment groups during the specified timeframe due to capacity constraints or other unforeseen hurdles	In advance of implementation roll out, the T4D study coordinators will work closely with the implementing partner to develop a sound operational plan outlining assumptions about the number of facilitators working with the communities; a plan for training facilitators; a staffing plan for managing and monitoring the facilitators; as well as a realistic timeline and plan for roll out of the intervention. The plan will be reflected in the contract and the associated budget agreed to before implementation. During implementation, the T4D study coordinators will monitor progress, identify any challenges and troubleshoot.

IV.3 - Estimating Program Impacts

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. Our estimation strategy consists of estimating the following regression equation:

$$(1) \quad Y_{ihjk} = \beta_0 + \beta_1 TREAT_{jk} + \mathbf{X}_k + \varepsilon_{ihjk}$$

In this equation, the variable Y_{ihjk} 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 household h in village j in catchment area k . The variable \mathbf{X}_k is a vector that includes all variables that were used to stratify prior to random assignment. The variable $TREAT_{jk}$ is an indicator variable that takes a value of 1 if the village was assigned to receive the treatment, and 0 otherwise. The coefficient β_1 provides the estimate of the impact of the program. Standard errors will be clustered at the village level using the standard Huber-White estimator to account for correlations in mother's or children's characteristics within villages.

Given that we plan to collect baseline data on households in the 200 villages and then collect follow-up data on another set of households within these same villages (repeated cross-section), we plan to estimate a second set of regressions that control for baseline characteristics at the *village* level (see equation 2). Note that these control variables are meant to help improve the statistical precision of our impact estimates, but we do not expect them to have a substantial effect on the magnitude of the impact estimates.

$$(2) \quad Y_{ihjk} = \beta_0 + \beta_1 TREAT_{jk} + \mathbf{X}_k + \beta_2 BASELINE_{jk} + \varepsilon_{ihjk}$$

IV.4 - Sub-group Analysis

We plan to estimate the impacts of the program on a number of key sub-groups. First, for analysis of birth in a facility and birth with a skilled attendant, whether a woman had previously given birth in a facility will be taken into consideration. We will look at three groups: women having their first child, women who have previously given birth only at home, and those who have given birth at least one time previously in a facility.

Second, for all other outcomes, we will look at village level characteristics that potentially affect the village's ability to act collectively. Specifically, we will look at community level 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).

Finally, we will consider the distance to the facility, since distance affects access to services as well as the kinds of actions communities may be likely to take. We will estimate separately impacts for those villages with a health facility in the village, those less than 3km away from the facility, and those between 3 and 7 km from a facility.

Being explicit about the sub-groups in advance is important to protect the research against conducting statistical tests *ex-post* and discovering spurious results. While we do not wish to discard the possibility of testing hypotheses that emerge from the implementation of the project and our qualitative work, we will be explicit about what hypotheses were specified at the outset and which ones arose after the design work. The final list of pre-specified sub-groups will be clearly stated in the pre-analysis plan.

To conduct sub-group analyses in the context of subgroups of only two categories (say male vs. female child), we add a sub-group indicator variable and sub-group-treatment interaction term as explanatory variables:

$$(3) Y_{ijk} = \beta_0 + \beta_1 TREAT_{jk} + \beta_2 SUBGROUP_{ijk} + \beta_3 SUBGROUP_{ijk} * TREAT_{jk} + X_k + \varepsilon_{ijk}$$

The coefficient β_3 on the interaction term provides the estimate of the difference in impacts between the sub-group that takes the value of 1 and the sub-group that takes the value of zero. For sub-groups with several categories, the procedure is similar except we would add sub-groups indicator variables for all categories except one (i.e. the reference or base group).

IV.5 - Statistical Power

We conducted statistical power calculations to determine the sample size needed to detect policy-relevant impacts. We used the Optimal Design software²³ to conduct our power calculations. We concluded that a sample size of 200 villages in each country, split equally into treatment and control groups, and each consisting of 30 households, would yield reasonable statistical power for our study. This section presents the key results of our power calculations and describes our assessment of the level of statistical power that they imply. We focused on 6 primary outcomes (see Table 1), using data from the most recent Demographic and Health Surveys (DHS) from Tanzania and Indonesia to perform our power calculations.

With standard parameters and assuming treatment villages in the catchment area of 100 facilities, and control villages in the catchment area of 100 facilities, Table 6 below illustrates the minimum detectable effect (MDE) sizes that our study design would be able to detect on key outcomes (values below which we would not be able to rule out a positive impact even if we did not detect one).

²³ Raudenbush, S. W., et al. (2011). Optimal Design Software for Multi-level and Longitudinal Research (Version 3.01). Available from www.wtgrantfoundation.org.

Table 5 - Minimum Detectable Effects for Primary Outcomes (Overall Sample)²⁴

Outcome	Tanzania (Tanga) ²⁵			Indonesia (South Sulawesi; Banten)		
	DHS 2012	MDE (δ)	Post-intervention value	DHS 2012	MDE (δ)	Post-intervention value
Facility births	41%	0.162	49%	48%; 61%	0.19	57%; 70%
Birth w/ skilled attendant ²⁶	45%	0.162	53%	83%	0.19	90%
Proportion of women receiving at least 4 ANC visits	39%	0.162	44%	83%	0.13	88%
Median months pregnant at first ANC visit	4.85	0.138	4.67	2.94; 2.12	0.16	2.67; 1.88
PCA Index of ANC components ²⁷	0.15	0.14	0.33	0; 0.166	0.16	0.19; 0.36
Mean weight- for-age score ²⁸	-1.2	0.09	-1.11	-	-	-

²⁴ Using data from Pagel et al. (2011), the highest estimated intra-cluster correlation used in our power calculations does not exceed 0.24. In DHS data from Tanzania, it does not exceed 0.025 and is typically <0.05 in most countries. Where reliable ICC data was not available, we assume an ICC of 0.10. For the other parameters in the power calculations we assume 200 clusters, each with 30 births. Standard levels for the power (0.80) and significance level (0.05 for a two-tailed test) are used as well. Finally, we assume that baseline measures will allow us to explain at least 10% of the variation in endline outcomes, thereby reducing the variance in outcomes of interest from the perspective of measuring the impact of the intervention.

²⁵ DHS 2012 data from the Tanga region of Tanzania was used in the power calculations, while corresponding data for the provinces of South Sulawesi and Banten were used in Indonesia. Where province-level data were unavailable in Indonesia, country-level data were used.

²⁶ We use the definition of a skilled attendant used by the DHS, defined as including “doctor/AMO, clinical officer, assistant clinical officer, nurse/midwife, and MCH aide.”

²⁷ In Tanzania, the following components of ANC care went into forming the PCA index: (1) whether anti-malarial tablets were taken during the pregnancy, (2) whether iron tablets/syrup were taken during the pregnancy, (3) whether blood pressure testing was done, (4) whether urine samples were taken, and (5) whether women were informed about complications with the pregnancy. Each of these is a simple yes/no question in the DHS survey. For each respondent, an index score is calculated based on a principal component analysis. In Indonesia, all the same components *except* for anti-malarial tablet intake went into construction of the index.

²⁸ The mean weight-for-age score is reported in standard deviations i.e. the z-score. There is no data available on the mean weight-for-age score in Indonesia.

These effect sizes are in line with some of the existing literature; for instance, Basinga et al. [2011] find a 23% increase in the number of institutional deliveries due to a pay-for-performance intervention in Rwanda, compared to minimum effects of 13-20% above. If the impact was at that magnitude, we would see 49% of deliveries in Tanzania and 56-69% in Indonesia take place in a health facility; given that these are rural rates, it is worth benchmarking them against urban rates in the two countries, which are at 83% and 96% respectively. The study is thus powered to detect even a modest closure of the rural-urban gap in facility births. Gertler and Vermeesch [2012] find that performance incentives in Rwanda increase weight-for-age by 0.53 standard deviations and height-for-age by 0.25 standard deviations; we are powered to detect even a modest 0.09 standard deviation decrease in the average weight-for-age score in Tanzania.

A variety of design options were considered before we arrived at the two-arm 200 facility design outlined above. For instance, we considered decreasing the number of clusters to either 100 or 150 facilities. A number of factors ultimately dictated the larger number of clusters in our design. First, given the open nature of the social action menu, it will be difficult to detect effects on outcomes associated with particular mechanisms unless we have a large overall sample size. Second, being powered to detect smaller effects than we anticipate sets us up to conduct better subgroup analyses, at both the household and cluster levels. For instance, our study is powered to conduct analysis by household terciles for effect sizes of less than 15%, well below levels found by Basinga et al. [2011] in Rwanda. Power calculations for sub-group analysis (at the tercile level) are presented in Table 6 and Table 7. Finally, one of our risk mitigation strategies is to be conservative in the statistical power calculations (see Table 4).

In addition, we also considered a design that would attempt to measure spillovers on non-treatment villages in a health facility's catchment area and a three-arm design that would include two different transparency and accountability treatments, but concluded that these designs would not have sufficient statistical power to detect policy-relevant impacts. A visual of the spillover design can be found in Appendix A.²⁹

²⁹ Many thanks to Theodore Svoronos for help with the spillover design.

Table 6 - Minimum Detectable Effects for Primary Outcomes Sub-Group Analysis of Household Characteristics (at the Tercile Level)

Outcome	Tanzania (Tanga) ³⁰			Indonesia (South Sulawesi; Banten)		
	DHS 2012	MDE (δ)	Post-intervention value	DHS 2012	MDE (δ)	Post-intervention value
Facility births	41%	0.19	51%	48%; 61%	0.19	57%; 70%
Birth w/ skilled attendant ³¹	45%	0.19	54%	83%	0.19	90%
Proportion of women receiving at least 4 ANC visits	39%	0.15	46%	83%	0.15	89%
Median months pregnant at first ANC visit	4.85	0.17	4.64	2.94; 2.12	0.17	2.66; 1.87
PCA Index of ANC components ³²	0.15	0.17	.22	0; 0.166	0.17	0.20; 0.49
Mean weight-for-age score ³³	-1.2	0.14	-1.06	-	-	-

³⁰ DHS 2012 data from the Tanga region of Tanzania was used in the power calculations, while corresponding data for the provinces of South Sulawesi and Banten were used in Indonesia. Where province-level data was unavailable in Indonesia, country-wide figures were used.

³¹ We use the definition of a skilled attendant used by the DHS, defined as including “doctor/AMO, clinical officer, assistant clinical officer, nurse/midwife, and MCH aide.”

³² In Tanzania, the following components of ANC care went into forming the PCA index: (1) Whether women took anti-malarial tablets during the pregnancy, (2) Whether iron tablets/syrup were taken during the pregnancy, (3) Whether blood pressure testing was done, (4) Whether urine samples were taken, and (5) Whether women were informed about complications with the pregnancy. Each of these is a simple yes/no question in the DHS survey. For each respondent, an index score is calculated based on a principal component analysis. In Indonesia, all the same components *except* for anti-malarial tablet intake went into construction of the index.

³³ The mean weight-for-age score is reported in standard deviations i.e. the z-score. There is no data available on the mean weight-for-age score in Indonesia.

Table 7 - Minimum Detectable Effects for Primary Outcomes Sub-group Analysis of Cluster (Village) Characteristics (at the Tercile Level)

Outcome	Tanzania (Tanga) ³⁴			Indonesia (South Sulawesi; Banten)		
	DHS 2012	MDE (δ)	Post-intervention value	DHS 2012	MDE (δ)	Post-intervention value
Facility births	41%	0.29	55%	48%; 61%	0.28	62%; 75%
Birth w/ skilled attendant ³⁵	45%	0.28	59%	83%	0.28	94%
Proportion of women receiving at least 4 ANC visits	39%	0.19	48%	83%	0.07	90%
Median months pregnant at first ANC visit	4.85	0.24	4.54	2.94; 2.12	0.24	2.53; 1.75
PCA Index of ANC components ³⁶	0.15	0.24	0.47	0; 0.166	0.24	0.29; .046
Mean weight-for-age score ³⁷	-1.2	0.16	-1.04	-	-	-

³⁴ DHS 2012 data from the Tanga region of Tanzania was used in the power calculations, while corresponding data for the provinces of South Sulawesi and Banten were used in Indonesia. Where province-level data was unavailable in Indonesia, country-wide figures were used.

³⁵ We use the definition of a skilled attendant used by the DHS, defined as including “doctor/AMO, clinical officer, assistant clinical officer, nurse/midwife, and MCH aide.”

³⁶ In Tanzania, the following components of ANC care went into forming the PCA index: (1) Whether women took anti-malarial tablets during the pregnancy, (2) Whether iron tablets/syrup were taken during the pregnancy, (3) Whether blood pressure testing was done, (4) Whether urine samples were taken, and (5) Whether women were informed about complications with the pregnancy. Each of these is a simple yes/no question in the DHS survey. For each respondent, an index score is calculated based on a principal component analysis. In Indonesia, all the same components *except* for anti-malarial tablet intake went into construction of the index.

³⁷ The mean weight-for-age score is reported in standard deviations i.e. the z-score. There is no data available on the mean weight-for-age score in Indonesia.

V – Case Study Research

The case study portion of our evaluation is primarily designed to explore several of our research questions that are either difficult or, given the current state of knowledge in the transparency and accountability field, infeasible to address with the randomized evaluation:

4. *Empowerment*: What is the effect of the intervention on citizen empowerment and efficacy, both perceived and actual?
5. *Mechanisms*: If there are significant effects, what are the mechanisms through which these effects occur?
6. *Context*: What is the role of context in shaping or determining these mechanisms?

Question 4: Empowerment

For our intervention to be successful and sustainable, it should ideally contribute to some degree of “community empowerment”—specifically, a recalibration of the relationship between the community and actors whose responsiveness and commitment is important to the efficient and effective functioning of health care. In addition to the primary outcomes for this question described above (which will be part of the RCT evaluation), assessing the degree to which our intervention contributes to this kind of “empowerment” requires understanding:

1. *The context of state-society relations*—including the characteristics of public institutions (e.g., mayors, local officials, local branches of state institutions) that interact with the community; accountability structures and their degree of effectiveness at the community, regional, and national level; the degree of political competition; perceptions around the degree to which individual providers, officials, and politicians are responsive to community concerns; as well as how, if at all, these relations change over the course of the intervention period;
2. *Configurations of local and traditional power*—including who takes responsibility for actions to improve, or hold accountable those responsible for improving, health care;
3. *Capacity for action, individual and/or collective*—including institutional and social structures that facilitate individual and/or collective action;
4. *Health care responsiveness*—both objective (instances of the system responding to community needs, mechanisms to provide system feedback) and subjective (community perceptions of the degree to which the system responds to their needs); and
5. *Community efficacy in general*—both objective and subjective.

As this question is about change, the key is to understand how these characteristics change, if they do, over the course of the intervention.

Question 5: Mechanisms

To assess mechanisms, we need to understand a number of features of how the information we present is understood and the actions, if any, that it triggers. The key elements are those of the “action cycle” developed by Fung, Graham, and Weil (2007). The hypothesis is that successful local governance Transparency and Accountability interventions trigger an action cycle with four main elements: 1) information *salient* to users triggers 2) *actions* by users in response to that information, to which 3) providers, officials, or other community members are *sensitive* and 4) *respond constructively*. To assess these mechanisms in the context of our intervention requires understanding:

1. *Information salience*—for example: Do participants in the intervention understand the information? Do they care? Do they perceive it as a problem? How do they come to understand it?
2. *Information usability*—for example: Does the information lead participants to want to change their behavior in some way? How do they figure out what to do with it? What actions, if any, are taken in response to the information?
3. *Target sensitivity*—for example: Do the targets of actions—e.g. providers, officials, or other community members—ignore those actions, or do they respond in some way?
4. *Target response*—for example: If the targets of action respond, is that response constructive or does it seek to subvert or undermine the actions? Does the response improve the situation, make it worse, or have no effect? Does the response set the stage for more or less constructive future engagement with the community?

Question 6: Context and its Interaction with Mechanisms

We expect mechanisms to differ with the context. In particular, to the extent that communities decide to engage with the health system rather than “self-help,” the “five worlds” framework we developed in the early phase of this project (Kosack and Fung 2014) describes two important dimensions of the political economy that we expect to fundamentally condition the kind of “action cycle” our intervention triggers: 1) the degree of competition around health care, which determines whether users have the exit option, and 2) the political environment, including a) the degree of accountability of both front-line health care providers and the officials and politicians who oversee them, and b) the inherent willingness of those providers, officials, and politicians to engage in reform efforts regardless of their institutional incentives and constraints. To assess these elements of the context requires understanding:

1. *The context of health care*—including health care and how it operates, including the degree of competition and prevailing prices for services;
2. *The context of each community*—including basic characteristics and local history around: community economic, social, and demographic characteristics; civil society presence and strength; and
3. *The context of state-society relations*—including all the elements described under Question 4.

The case study portion of the evaluation relies on a number of interconnected data collection approaches: questions in surveys and community focus groups administered to all treatment and control communities; facilitator reports and social action plans from all treatment communities; coding of the scorecard, social action, and 90-day follow-up meetings using a standard coding scheme (SCS) and seating chart, paired with a short empowerment survey; and deep ethnographic analysis of a sample of four treatment and two control communities in each of the two sites. These data collection approaches will be used across four samples, or “data collection layers,” from among our 200 treatment and control communities per site. In order of breadth of data collection, the four layers are:

1. All treatment and control communities: 1) survey questions on the RCT surveys and community focus groups run in conjunction with the RCT and 2) facilitator reports/social action plans (treatment communities only).
2. Subsample of 40 treatment communities: 1) coded community meetings (scorecard, social action, and the 90-day follow-up meeting) using a standard coding scheme and 2) key informant interviews at 90 days.
3. Subsample of 20 treatment communities: 1) coded meeting attendee participation through use of a seating chart and 2) an empowerment survey.
4. Subsample of 4 treatment and 2 control communities: field work by ethnographers (1-2 months before the intervention, 4-5 months during the intervention, and 1-2 months after the intervention).

Table 8 outlines the data collection samples and methods that will be used to gather information on the sub-questions for each of the four main research questions described above. (The layers and their integration are described in more detail below in Section V.2 and in Figure 5.)

Table 8 - Research Questions, Samples, and Methods

Research Question	Sub-questions	Sample Layers	Data Collection Methods (Layers)
4. Empowerment	1. Beliefs and practices of community members	1, 2, 4	<ul style="list-style-type: none"> ▪ RCT surveys/focus groups (1) ▪ Coding on a standard coding scheme/key informant interviews (2) ▪ Ethnography (4) ▪ Facilitator reports (1)
	2. Beliefs and practices of providers and officials	1, 2, 4	<ul style="list-style-type: none"> ▪ RCT surveys/focus groups (1) ▪ Coding on a standard coding scheme/key informant interviews (2) ▪ Ethnography (4)
	3. Perceptions of power and efficacy	1, 2, 3, 4	<ul style="list-style-type: none"> ▪ RCT surveys/focus groups (1) ▪ Coding on a standard coding scheme/key informant interviews (2) ▪ Empowerment survey (3) ▪ Ethnography (4)
	4. Distribution of influence and action across the intervention	2, 3, 4	<ul style="list-style-type: none"> ▪ Coding on a standard coding scheme/key informant interviews (2) ▪ Participation coding (3) ▪ Ethnography (4)
5. Mechanisms	1. Participation in the intervention meetings	1, 2, 3, 4	<ul style="list-style-type: none"> ▪ Coding on a standard coding scheme/key informant interviews (2) ▪ Participation coding (3) ▪ Ethnography (4) ▪ Facilitator reports (1)
	2. Attendees' understanding of the information presented (information salience)	2, 4	<ul style="list-style-type: none"> ▪ Coding on a standard coding scheme/key informant interviews (2) ▪ Ethnography (4)
	3. Attendees' understanding of the relationship of the information to a health problem (quality of agenda setting)	1, 2, 4	<ul style="list-style-type: none"> ▪ Coding on a standard coding scheme/key informant interviews (2) ▪ Ethnography (4) ▪ Facilitator reports (1)
	4. Attendees' planning for what to do with the information (quality of the planning)	1, 2, 3, 4	<ul style="list-style-type: none"> ▪ Coding on a standard coding scheme/key informant interviews (2) ▪ Participation coding (3) ▪ Ethnography (4) ▪ Facilitator reports/social action plans (1)
	5. Actions undertaken, individual or collective, long or short route	1, 2, 3, 4	<ul style="list-style-type: none"> ▪ RCT surveys/focus groups (1) ▪ Coding on a standard coding scheme/key informant interviews (2) ▪ Participation coding (3) ▪ Social action plans (1) ▪ Ethnography (4)
	6. Provider or official sensitivity to actions	1, 2, 4	<ul style="list-style-type: none"> ▪ RCT surveys/focus groups (1) ▪ Coding on a standard coding scheme/key informant interviews (2) ▪ Ethnography (4)

Research Question	Sub-questions	Sample Layers	Data Collection Methods (Layers)
			<ul style="list-style-type: none"> ▪ Social action plans (1)
	7. Constructiveness of provider or official response	1, 2, 4	<ul style="list-style-type: none"> ▪ RCT surveys/focus groups (1) ▪ Coding on a standard coding scheme/key informant interviews (2) ▪ Ethnography (4) ▪ Social action plans (1)
	8. Evolution of community actions based on response	1, 2, 4	<ul style="list-style-type: none"> ▪ RCT surveys/focus groups (1) ▪ Coding on a standard coding scheme/key informant interviews (2) ▪ Social action plans (1) ▪ Ethnography (4)
	9. Actions taken that have no plausible connection to the intervention	1, 2, 4	<ul style="list-style-type: none"> ▪ RCT surveys/focus groups (1) ▪ Coding on a standard coding scheme/key informant interviews (2) ▪ Social action plans (1) ▪ Ethnography (4)
6. Context	1. Community Characteristics	1, 2, 4	<ul style="list-style-type: none"> ▪ RCT surveys/focus groups (1) ▪ Key informant interviews (2) ▪ Ethnography (4) ▪ Facilitator reports (1)
	2. Health care characteristics	1, 2, 4	<ul style="list-style-type: none"> ▪ RCT surveys/focus groups (1) ▪ Key informant interviews (2) ▪ Ethnography (4) ▪ Facilitator reports (1)
	3. Civil society	1, 2, 4	<ul style="list-style-type: none"> ▪ RCT surveys/focus groups (1) ▪ Key informant interviews (2) ▪ Ethnography (4) ▪ Facilitator reports (1)
	4. State institutions	1, 2, 4	<ul style="list-style-type: none"> ▪ RCT surveys/focus groups (1) ▪ Key informant interviews (2) ▪ Ethnography (4) ▪ Facilitator reports (1)
	5. Configurations of local and traditional power	1, 2, 4	<ul style="list-style-type: none"> ▪ RCT surveys/focus groups (1) ▪ Ethnography (4) ▪ Facilitator reports (1)
	6. Capacity for action, individual and collective	1, 2, 3, 4	<ul style="list-style-type: none"> ▪ RCT surveys/focus groups (1) ▪ Key informant interviews (2) ▪ Participation coding (3) ▪ Ethnography (4)
	7. Trust and solidarity	1, 2, 4	<ul style="list-style-type: none"> ▪ RCT surveys/focus groups (1) ▪ Key informant interviews (2) ▪ Ethnography (4)
	8. Social cohesion and inclusion	1, 2, 4	<ul style="list-style-type: none"> ▪ RCT surveys/focus groups (1) ▪ Key informant interviews (2) ▪ Ethnography (4)

V.1 – Case Study Design Orientation

Our case study methods are intended to be both deductive and inductive in nature. We have tried to be explicit about our theoretical priors in the “five worlds” theoretical framework that guides this project (Kosack and Fung 2014), so as to gather information on the specific elements necessary to understanding how the intervention plays out in the

treatment communities. But we have also tried to design the case study elements of our research to allow for a range of possible families of explanations, including those that we could not have anticipated prior to the intervention. In particular, we have tried to develop an approach that allows for both “rational choice”—or “substantialist”—explanations as well as more complex, contingent, relational explanations that may emerge from our focused investigations (De Renzio 2006, Ranson 2003, Mercer 2002, Eyben 2010). We will examine our data for similarities and differences in mechanisms based on a political economy logic of interaction among groups of boundedly rational actors—such as various groups of political elites, state agencies, and groups of citizens distinguished along various demographic dimensions. At the same time, we intend to use direct observation and ethnographic techniques in order to identify emergent mechanisms, groups, and conditions, which a rational-actor model might miss (Scott 1987, 1999, Joshi and Houtzager 2012). Our object is to approach our case work with the initial hypotheses necessary for rigorous research (King, Keohane, and Verba 1994) while avoiding as much as possible preconceived notions of mechanisms or context by allowing for the exploration of factors, issues, and hypotheses generated in the field.

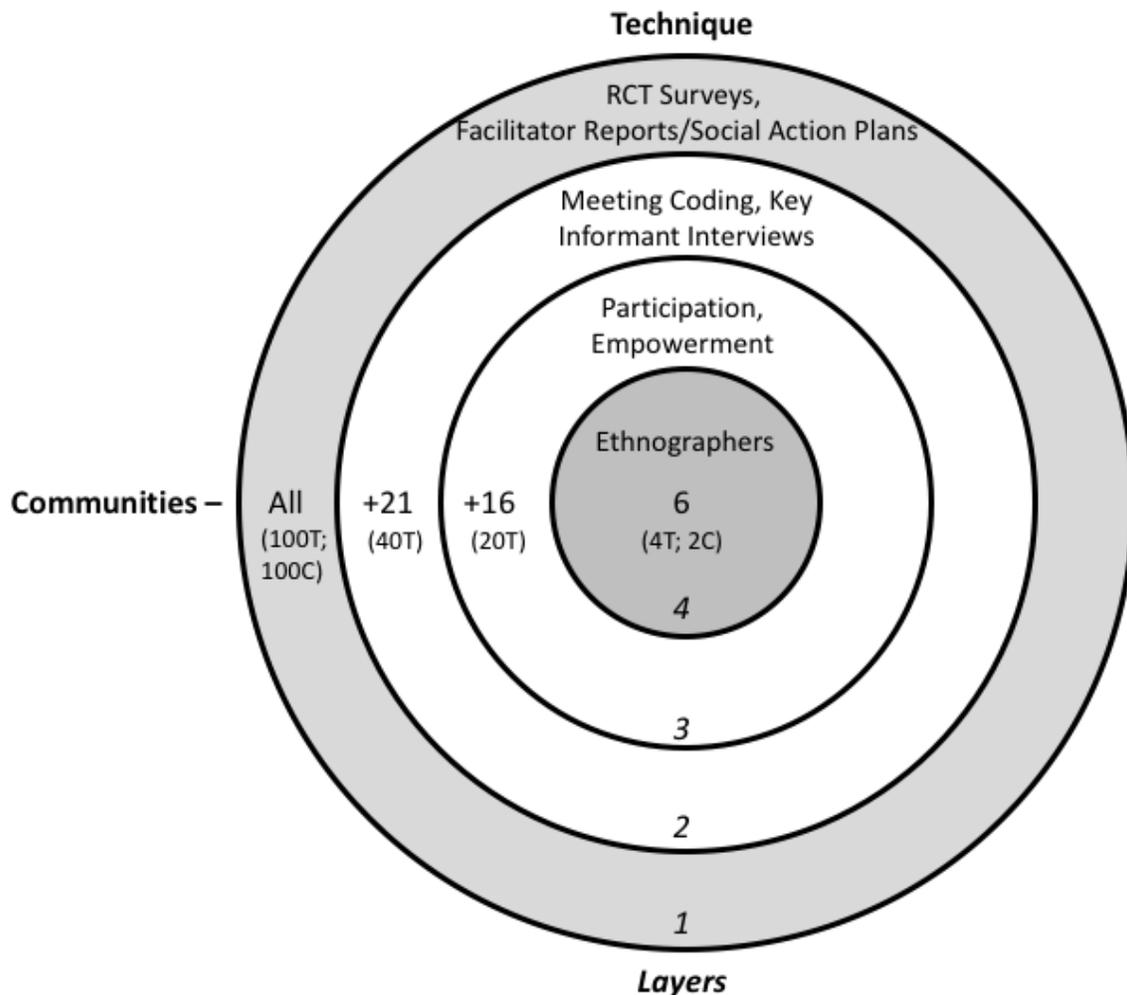
V.2 – Case Study Data Collection Layers and Integration with RCT Evaluation

As mentioned in the previous section, the case study data collection will have four layers, integrated with the RCT evaluation. Each layer is designed with two purposes: 1) to provide different kinds of information about communities and the processes triggered by our intervention, and 2) to check the information of the other layers.

This multi-layered approach is intended to allow for triangulation and minimization of the bias inherent in any one data-collection method (retrospective bias, Hawthorne effects, etc.): data collected in one layer can be checked against and combined with data collected in other layers to provide more robust information about each of the key research questions. For example, in determining how communities decided on social action, we will be able to draw on information from facilitator reports and social action plans in all treatment communities, coding of the social action meetings on a standard coding scheme in 40% of the treatment communities, key informant interviews, and ethnography in 4 of the treatment communities (2 of which ethnographers will live in, 2 of which they will visit regularly). On their own, each of these data collection methods has myriad biases and other flaws; together, however, they can check each other, providing information about the process by which communities of different kinds and with different characteristics decide whether and how to act on the information provided by our intervention.

Figure 5 shows the four layers, the number of communities in each, and the primary data-collection agent for each. The remainder of this section further explains the basic data collection strategy in each layer; the specific methods and the elements that we will attempt to assess in each layer are described in Chapter VI.

Figure 5 - Layers of Data Collection (the T4D Onion)



Notes: “T”—Treatment Communities; “C”—Control Communities.

A. *Layer 1: Surveys of All Communities.* We will collect a range of data alongside the baseline and endline surveys collected for the RCT. Survey questions will be structured around our five worlds framework and the contextual factors relevant to it, as well as other contextual factors that might potentially be relevant to how our intervention plays out (e.g. education levels, economic characteristics, ethnic and social cleavages, etc.). There are several purposes of this element of the data collection. First, it will provide standard data for all communities. Second, it will provide a check against the standard coding scheme used in the other layers, since the survey questions and the standard coding scheme will have overlapping elements. Third, it will provide a check against the Hawthorne effect, as the data collection in the other layers all rely on observers (that is, are there differences in seemingly similar communities when one has an observer?).

In addition to household surveys, the RCT will include a baseline and endline community focus group (CFG) in each treatment and control community. For the case study evaluation, these CFGs will enable us to gather data on village-wide variables (such as the existence of specific community groups and events that household survey participants themselves may not be involved in). The CFGs will be used to record all active community groups (e.g. civic, religious, informal, etc.). They will also be used to gather specific context variables, such as details on the village hierarchy and existence of donor programs. In the endline, the community focus group protocol will be expanded to include questions on the specific community actions taken as a result of the T4D intervention.

Separately from the RCT surveys, we will compile information gathered by facilitators on community context, the intervention, and any planned social action that results. This information will provide supplementary data on community context and the implementation of the intervention, as well as data on the social action plans that each community developed. It will also include information on revisions to the social action plans.

B. Layer 2: 40 Treatment Communities. Additional data will be collected for a subset of communities, in which we will observe intervention meetings, using a standard coding scheme (described below), to gain a basic understanding of how communities understand the information presented to them and what they do with it, both to diagnose problems with health care and actions they decide to take to fix those problems. Elements of the coding scheme will focus on our “five worlds” framework as well as features of the community context. The purpose of data collection with this standard coding scheme—which will be used in all Layers 2-4—is to provide a range of standard data on the community and its use of the intervention for use in large-N inductive analyses (described below). As part of this process, we will also conduct key informant interviews in this subset of communities. The key informant interviews will provide a “snapshot” of what the community has attempted and accomplished approximately 90 days after the social action planning meeting. The CSO facilitator, community representatives, village leadership, health professionals, and specific targets of the social actions will serve as key informants.

C. Layer 3: 20 Treatment Communities. In a subset of 20 of the Layer 2 treatment communities, a separate observer will code meeting attendee participation through use of a seating chart, on which the observer will mark, for each participant, meeting attendance times, movement throughout the day, times they spoke distinctly from the rest of the group and instances of volunteering to do an activity, being assigned to do an activity, and reporting having done an activity. The purpose of this protocol is to understand the kind of participation of community representatives, ranging from unengaged to engaged in the discussion and the activities that emerge. In addition, each participant will be given a short survey about their sense of empowerment and efficacy once before the scorecard meeting and a second time after the 90-day follow up meeting. This survey will be combined with the participation tool to determine if the nature or degree of

participation in the intervention is related to any changes in CRs' perceptions of their empowerment.

D. Layer 4:4 Treatment and 2 Control Communities. Additional data will be collected by ethnographic researchers over a period of 6-8 months in which we will try to gain a deeper understanding of the existing context, and, in the treatment communities, the implementation of the intervention, and its implications for health outcomes and the broader political and social structure. The same coding scheme used in Layer 2 will be used in these communities. In addition, the data collection will focus on the nature and organization of the existing health care system at baseline and over the course of the intervention, beliefs and practices of providers—including the perceptions and biases behind their practices—and power relationships between and among providers, officials, and community members. Case studies will rely both on subjective assessments of providers, officials, and community members, as well as objective data that can be gathered from health facilities (who was treated, how, reactions to complaints, etc.) and community meetings (who sat where, who spoke). Case studies will also concentrate on “emergent” sub-groups and mechanisms, which can then be incorporated into endline surveys. The purpose of adding this ethnographic component is to collect information that the standard coding scheme misses, including information that it misses about how communities understand, use, and act on the information provided in the intervention.

VI - Sampling and Data Collection Plans

The study will involve collecting data for 200 villages in each of the two countries. Table 9 summarizes the key elements of the data collection.

Table 9 - Data Collection Plan

Data Collection	Target Sample	Key Goals
Baseline Household Survey	Random sample of 15 mothers in Tanzania, 12 mothers in Indonesia who have given birth in the last year in each of the 200 study villages	<ul style="list-style-type: none"> • Verify baseline equivalence between treatment and control villages • Create baseline measures to increase statistical power in impact estimation • Collect data on variables for sub-group analyses
Baseline Facility Survey	200 facilities that are part of the study (Tanzania and Indonesia) plus other facilities that provide services to the 200 villages that are part of this study (Indonesia only)	<ul style="list-style-type: none"> • Verify baseline equivalence between treatment and control villages • Collect data on variables for sub-group analyses
Baseline Community Focus Groups	Key community informants, including village chief, informal leader(s) and community health volunteers, in 200 study villages	<ul style="list-style-type: none"> • Collect data on variables for sub-group analyses • Collect background data on community context.
Facilitator Assessments/Social Action Plans	All facilitator-led meetings in the 100 treatment villages	<ul style="list-style-type: none"> • Information on the social action plans and the evolution of those plans • Provide a secondary data source on the community context and on the intervention
Coding of meetings on a Standard Coding Scheme (CSC)	Scorecard, social action, and follow-up meetings in 40 treatment communities	<p>Gain a basic understanding of:</p> <ul style="list-style-type: none"> • How communities understand the information presented to them • What they do with that information, both to diagnose problems with health care and actions they decide to take to fix those problems
Key Informant Interviews	CSO facilitators, community representatives, village leadership, health professionals, and specific	<ul style="list-style-type: none"> • Verify the social actions as reported in the social action plans

Data Collection	Target Sample	Key Goals
	targets of the social actions in 40 treatment communities (same as CSC)	<ul style="list-style-type: none"> Gain additional details on the social actions, including responses of providers or officials and any adaptation or course-correction that the community undertakes
Participation Monitoring	Scorecard, social action, and follow-up meetings in 20 treatment communities (subset of the CSC communities)	<ul style="list-style-type: none"> Understand the kind of participation of community representatives, ranging from unengaged to engaged in the discussion and the activities that emerge
Empowerment Surveys	All community representatives in the 20 treatment communities with participation monitoring (up to 16 attendees per community)	<ul style="list-style-type: none"> Determine changes in CRs' perceptions of their empowerment
Ethnography	4 treatment and 2 control communities, over 6-8 months (1-2 months before the intervention, 4-5 months during the intervention, and 1-2 months after the intervention)	<ul style="list-style-type: none"> Gain a deeper understanding of the existing context In the treatment communities, uncover the implications of the intervention for health outcomes and the broader political and social structure Explore unintended experiences and consequences of the program
Endline Household Survey	Random sample of 30 mothers who have given birth in the last year in each of the 200 study villages	<ul style="list-style-type: none"> Estimate impacts of intervention
Endline Facility Survey	200 facilities that are part of the study plus other facilities that provide services to the 200 villages that are part of this study	<ul style="list-style-type: none"> Estimate impacts of intervention
Endline Community Focus Groups	Key community informants, including village chief, youth leader, religious leader(s) and other active community members, in 200 study villages	<ul style="list-style-type: none"> Collect information on social actions and the outcomes of these actions

This section specifies the sampling plan that will be used in the impact evaluation and qualitative case-study work, and the corresponding data collection plan. We begin by describing the sampling frame for the project's data collection. We then proceed to describe our data collection strategy and the content of the data collection instruments.

VI.1 - Sampling Frames

Household Surveys: To develop the village-level household sampling frame, we will conduct a complete census of all households in a village and identify those with women who gave birth in the last year. We will then randomly select households (12-15 at baseline and 30 at follow-up) from each village to be surveyed.

Facility Surveys: In each country, we will conduct a facility survey in the 200 sampled facilities that define the village catchment area (see Section IV.2 - Randomization Design and Implementation).

Community Focus Groups: A community focus group will be conducted in the same villages as the household surveys. These villages are randomly selected as described in Section IV.2. The survey firm will select community focus group participants at the same time they sample households. Selection will not be random and participants will vary based on who is available. Targeted participants include: the village executive officer, youth leaders, religious leaders, representatives from women's groups, and other active community members.

Facilitator Reports, including Social Action Plans: The CSO facilitators will submit reports on each village in which they work (all treatment communities). These reports include four versions of the social action plans: the original social action plan and a revised social action plan from each of the 30-, 60- and 90-day follow up meetings.

Standard Coding Scheme (SCS) and Key Informant Interviews: Meeting coders will observe and code three meetings (scorecard, social action, and the final follow-up meeting) and conduct key informant interviews in each of 40 treatment communities. This pool of treatment communities includes those that are covered by the ethnographers. The communities will not be randomly selected, but will be geographically diverse and distributed amongst all treatment phases.³⁸

Participation Monitoring and Empowerment Survey: Meeting coders will observe participation in meetings (using a drawing) in a subset 20 of the SCS/KII communities. These communities do not include those covered by the ethnographers. The empowerment survey will be conducted in these same communities. The baseline empowerment survey will be administered to all CRs before the start of the community scorecard meeting. The endline empowerment survey will be administered to the same participants upon the conclusion of the 90-day follow up meeting.

³⁸ In Indonesia, the treatment was spread over two phases. In Tanzania, the treatment was spread over four phases. The number of phases was logistical in nature and determined in conjunction with the CSO partners.

Ethnography: Short-term ethnography will be conducted by two researchers in a total of four treatment and two control communities per country. Sampling will not be random, since we need to consider geographic proximity (each ethnographer will cover three communities—two treatment and one control). Sampling will be stratified by province (Indonesia) or region (Tanzania), so that there is one researcher per province/region. Finally, since a multitude of local languages are spoken across South Sulawesi and Banten (Indonesia), we will take into account the language abilities of the ethnographers when choosing placement. Depending on project resources and operational feasibility, we may also try to allocate additional ethnographic researchers to communities that are either 1) facing significant challenges in their social actions, 2) seeing significant success, or 3) changing strategy or tactics substantially when expectations about what social actions would be successful turned out not to be accurate.

VI.2 - Data Collection Plan

As described in the previous chapter, our data collection involves 4 layers (see the “T4D Onion” in

Figure 5). We will collect baseline and endline data from 200 villages and 200+ dispensaries Tanzania and 200 villages and 200+ puskesmas in Indonesia, as well as a variety of data from smaller samples of communities as part of the case study research.

Baseline: Tanzania (Layer 1—All Communities)

In Tanzania, the baseline will consist of three parts: a facility survey, a household survey and a community focus group. Data collection will commence in early 2015.

Facility surveys will be conducted in 200+ health dispensaries (see Section VI.1 - Sampling Frames). The survey will involve 1) collecting facility level information and 2) conducting interviews with key staff. It is expected that each facility survey will last 2-4 hours.

Household surveys will be conducted in one randomly selected village from each dispensary’s catchment area. The interview teams will interview 15 households per village.³⁹ These households will be randomly selected from all eligible households in the village. To be eligible, the household must include one woman who has given birth in the past year, and the survey will be administered to this woman. If the household includes more than one woman who gave birth in the past year, the respondent will be randomly selected from all eligible respondents.

Baseline: Indonesia (Layer 1—All Communities)

³⁹ We had initially proposed interviewing 30 households per village in the baseline. Since we have since decided on a repeated cross-sectional design, interviewing 30 households is no longer necessary. Twelve to sixteen households per village will give us enough information to verify random assignment. Thirty households per village will be interviewed in the endline.

As in Tanzania, the baseline in Indonesia will consist of a facility survey, a household survey, and community focus groups. The baseline survey will also begin in early 2015.

The facility surveys will be conducted in a total of 200+ puskesmas (see Section VI.1 - Sampling Frames). The survey will involve 1) collecting facility level information and 2) conducting interviews with key staff. It is expected that each facility survey will last 2-4 hours.

Household surveys will be conducted in one randomly selected village from each puskesmas's catchment area. The interview teams will interview 12⁴⁰ households per village. These households will be randomly selected from all eligible households in the village. To be eligible, the household must include one woman who has given birth in the past year, and the survey will be administered to this woman. If the household includes more than one woman who gave birth in the past year, the respondent will be randomly selected from all eligible respondents.

Household Survey Eligibility: Because our primary outcomes center on maternal and neonatal health, to be eligible, a household must include at least one woman who gave birth in the previous 12 months. See Section VI.1 (Sampling Frames) for a detailed explanation on sampling strategy.

Standard Coding Scheme and Key Informant Interviews (Layer 2)

The primary data-collection method will be objective and subjective coding of core aspects of the scorecard meeting, social action meeting, and the 90-day follow up meeting, using a standard coding scheme designed to assess the community context and aspects of how communities understand and use the information in the intervention. The specific aspects are described in Section VI.4 (Case Study Descriptive Elements). The coding scheme will include specific definitions of variables and standard procedures for assessment. Definitions will be included in a codebook that specifically describes the definition of each variable and how to observe each value of it.

Assessment will be in two phases. First, local researchers from the survey firm will observe the meetings. During observation, the researcher (known as a 'coder') will answer a set of questions for each part of the meeting designed to assess the discussion: how communities understood the information, how they came to a decision about the levers and barriers to focus on, how they decided on social actions, etc.

In addition to meeting observation and coding, the local researchers will conduct a series of key informant interviews in conjunction with the follow up meeting. These interviews will take place the day after the meeting. The purpose of these interviews is to assess what actions were taken, the audiences to which the actions were directed, whether the audience was receptive, whether the action(s) worked, why or why not, and community satisfaction with the chosen actions and how these actions were carried out. Target respondents include: the CSO facilitator, community representatives and other

⁴⁰ Ibid.

community members who were actively involved in the action(s), a representative from the health facility, the village head, and other actors who were targeted as part of the intervention.

Participation Monitoring and Empowerment Surveys (Layer 3)

Participation monitoring will take place in 20 of the Layer 2 communities. To observe participation, a second coder will be present in the same set of three meetings (scorecard, social action, and 90-day follow up). This coder is responsible for producing a drawn seating chart of meeting participants, including the CRs, the facilitators, and any additional participants. The coder will mark, for each participant, meeting attendance times, movement throughout the room, times they spoke distinctly from the rest of the group and instances of volunteering to do an activity, being assigned to do an activity, and reporting having done an activity.

In addition, each CR will be given a short survey about their sense of empowerment and efficacy once before the scorecard meeting (administered the day before or the morning of) and a second time after the 90-day follow up meeting (administered immediately after or the next day). This survey will be combined with the participation tool to determine if the nature or degree of participation in the intervention is related to any changes in CRs' perceptions of their empowerment.

Ethnographic Data Collection (Layer 4)

Ethnographic data-gathering will be the responsibility of a team ethnographic researchers on short term assignments. The inductive nature of the ethnographic approach in these communities requires some flexibility in the particular methods. Prior to entering the field, the ethnographic researchers will meet for several days as a group to jointly develop specific questions to explore in the categories described below. These preliminary questions will serve to focus the data collection efforts within each village and to organize the data collection across villages to facilitate comparison. Individual researchers will also be encouraged to explore additional questions and hypotheses that they develop in the field. Given the nature of ethnographic work, the specific techniques these ethnographers use will vary with the community and the question, but we anticipate relying on two types of methodologies:

1. Participatory rural appraisal techniques, such as:
 - a. Open meetings (an initial open meeting around the start of the intervention, a meeting at the end of the intervention, and a follow-up meeting several months after the intervention)
 - b. Interviews
 - c. Focus group discussions
 - d. Preference ranking
 - e. Maps and modeling using local analysts, such as Transect Walks to understand the location and distribution of health resources and usage in the village (what resources are available, what is easily accessible, to

- whom, where do people go when they cannot get access to their first choice for care)
- f. Stakeholder analysis of persons, groups, organizations, institutions, organized around enablers and saboteurs, and focusing around 1) each stakeholder's impact—the role the stakeholder needs to play if the intervention is guaranteed to be successful; the likelihood that they will play that role; and the consequences of them not playing that role—and 2) actions that can be taken so the stakeholder adds value
 - g. Venn Diagrams, for example to assess power and influence (size of circle is power or influence; overlaps are commonality or mutual influence)
 - h. Gender and other subgroup analysis
 - i. Wealth and resource ranking to assess relative economic standing, in which participants brainstorm indicators of wealth and resources and then categorize households by how well they are endowed on these indicators
 - j. Seasonal and historical diagramming
 - k. Immediate and continuous report-writing and reflections
2. *Direct Observation* of the intervention meetings, including of:
 - a. Who speaks, who seeks to be responsive and to whom are they trying to be responsive;
 - b. Spatial configuration (e.g. where are people sitting)
 - c. What problems are raised (e.g. who proposes a problem, who gets blamed)
 - d. The response to the problem, including 1) the initial response, including who that response serves and who takes on the responsibility for implementing it, and 2) the implementation of the response, including which sub-groups shoulder the burden and who gets the credit or blame when things go well or poorly

Facilitator Reports (Layers 1—All Treatment Communities)

In the course of implementing the intervention, facilitators (supervised by CSO partners) will gather information on the community context and on the intervention—such as who attended the meetings, community interest in and discussion of the information, and planning around social action, including copies of the social action plans at four time points. This information will be aggregated to provide a secondary data source on communities and the implementation of the intervention. Additionally, the social action plans will be coded and analyzed.

Endline (Layers 1—All Communities)

The endline will consist of three parts: a facility survey, a household survey and a community focus group. The baseline survey will begin 22 months after the completion of the intervention and will be administered over a period of approximately three months.

In Tanzania, the facility survey will be conducted in the same 200+ health dispensaries (and other health facilities) as the baseline. In Indonesia, the facility survey will be conducted in the same 200+ puskesmas (and potentially additional health facilities). The

survey itself will include a module on activities related to the intervention, but will otherwise be similar to the baseline questionnaire.

As described above, we have elected to use a repeated cross-sectional design for the household survey. We plan to interview 30 households in each of the same 400 villages (200 in Tanzania and 200 in Indonesia) as the baseline. These households will be randomly selected from all eligible households in the village (using the same strategy as the baseline, as described in Section VI.1) and may or may not be the same households interviewed in the baseline. The household questionnaire will be similar to the survey used in the baseline.

Finally, we will conduct an endline community focus group in each village. The instrument will be similar to the one used during the baseline, with the exception of an additional module on activities related to the intervention.

VI.3 – Survey Instruments

We employ three types of baseline/endline surveys in our study: 1) household, 2) facility, and 3) community focus group. Each survey will be administered twice – once before the intervention is initiated in the community (baseline), and again approximately two years later (endline). More details on the operationalization of the data collection are detailed in Section VI.2 (Data Collection Plan).

Household Survey (Layer 1—All Communities)

The household survey will consist of two parts, an interview with a woman who has given birth in the past 1 year and anthropomorphic measurements of the woman's most recent child.

The household survey will be used to measure primary outcomes. The primary outcomes and associated sample survey questions/anthropomorphic measurements are summarized in Table 10.

Table 10 - Sample Survey Questions⁴¹

Outcome	Questions
1. Birth in a health facility	Where did you give birth to [NAME]?
	What is the name and location of the place you delivered [NAME]?
2. Birth with a skilled provider	Who provided care during [NAME]'s birth? Check all that apply.
	What is the name of the person who provided care?
3. Four (or more) ANC visits	During the pregnancy did you ever have a pregnancy check-up?
	Where did you go for pregnancy check-ups?
	What is the name and location of the provider you visited?
	How many visits did you make for pregnancy check-ups?
4. First ANC visit within 3 months of pregnancy	How many months pregnant were you when you went for a pregnancy check-up?
5. Postpartum care within 7 days of delivery (mother)	Since you have given birth, has anyone done any of the following to check on your health?
	Did this check take place within seven days (one week) of giving birth to [NAME]?
6. Postnatal care within 7 days of delivery (baby)	In the time since [NAME] was born, did anyone do any of the following to check on [NAME]'s health?
	Did this check take place within seven days (one week) of giving birth to [NAME]?
8. Weight-for-age (z-score)	Weight in kilograms
	Date of birth
9. Citizen Empowerment	In the past 12 months, have you or a household member participated in communal activities where people came together to work for the benefit of the community?
	In the past 12 months, have you done any of the following? A. Attend a village/neighborhood council meeting, public hearing, or public discussion group B. Met with a politician, called him/her, or sent a letter C. Participated in a protest or demonstration D. Participated in an information or election campaign E. Alerted newspaper, radio or TV to a local problem F. Notified police or court about a local problem
	Are you or is someone in your household a member of any groups, organizations, or associations? Which groups?
	Do you consider yourself/household member to be active in the group, such as by attending meetings or volunteering your time in other ways, or are you relatively inactive? Are you/household member a leader in the group?
	How much of the time do you think Members of Parliament try their best to listen to what people like you have to say?
	How much of the time do you think Local Government Officials try their best to listen to what people like you have to say?
	If members of this community had a complaint about this health facility and brought it to the attention of the provider, do you think s/he would try to make an improvement?
	Do you feel that you have the power to make important decisions that can change the course of your life? Rate yourself on a 1 to 4 scale, where 1 means being totally unable to change your life, and 4 means having full control over your life.

⁴¹ Modeled after questions from the DHS6, the Indonesia Family Life Survey (IFLS4), (Grootaert et al. 2004), and the Social Economic Survey of Indonesian Households.

In addition to measuring primary outcomes, the household survey will measure a number of secondary outcomes. The survey tool itself consists of the following modules:

- The *household characteristics* module will include questions on the gender and number of household members, age, ethnicity, and education level.
- The *household socio-economic status* module will be used in conjunction with the household characteristics module to estimate socioeconomic status of the household.
- The *birth history* module captures all pregnancies, stillbirths, live births, and deaths, and enables us to estimate the infant mortality rate.
- The *health knowledge* module measures respondents' knowledge of safe antenatal, birth, and postnatal practices and behaviors.
- The *health seeking behavior and practices* module will capture the same indicators presented on the community scorecard⁴². This module serves two main purposes: 1) it allows us to verify the accuracy of the data collected by the community and facilitator, and 2) it allows us to measure the specific indicators presented in the scorecard.
- The *health experiences and perceptions* module enables us to measure the respondents' experiences with the formal and informal healthcare system and their perceptions of these systems.
- The *empowerment and efficacy* module will assess three indicators of empowerment: 1) general empowerment, such as control over decisions and the power to improve one's life, 2) perception of local government to be honest and responsive, and 3) levels of social capital (networks, trust, social cohesion and inclusion).
- The *depression* module will assess whether the mother meets standard criteria for depression.

Facility Survey (Layer 1—All Communities)

The facility survey will be used to assess the conditions at the facility. It will also be used to gain an understanding of services provided and utilization rates.

Data will be collected in three ways. Observation will be used to determine the facility condition: cleanliness, drug stocks, presence of key equipment, and size. Key staff interviews will be used to assess staff knowledge about maternal, neonatal and infant care, and to gather data on services provided and human resources at the facility. Finally, the facility records themselves will be analyzed to determine the number of patients seeking maternal and neonatal health services, and verify the catchment area the facility serves. These records will also be checked for completeness.

⁴² See Table 1 for a list of information presented on the scorecard. Note that some of the information in Table 1 will be collected using a facility survey and will not be included in the household survey.

Community Focus Groups (Layer 1—All Communities)

Community focus groups will enable us to gather data on village-wide variables (such as the existence of civic groups and events that household survey participants themselves may not be involved in).

The community focus group will be used to record all active community groups (e.g. civic, religious, informal, etc.). It will also be used to gather specific context variables, such as details on the village hierarchy and existence of donor programs. In the endline, the community focus group questionnaire will be expanded to include questions on the specific community actions taken as a result of the T4D intervention.

VI.4 – Case Study Descriptive Elements

The case study portion of our evaluation will seek to understand a variety of additional elements. Because one of the purposes of organizing the data collection into the four layers is to allow information gathered in one layer or with one methodology to check that gathered in another layer or with another methodology, there is by design a great deal of repetition in the elements below.

Standard Coding Scheme (Layer 2).

The standard coding scheme will include assessment of a targeted set of elements around the scorecard meeting, social action meeting, and 90-day follow up meeting:

Mechanisms, emergent and predicted (Question 5). Coders will record the mechanisms, if any, the intervention triggered.

During the scorecard meeting, this includes:

1. Who (what demographic groups) attended meetings.
2. How CRs and community members understood the information (information and its salience):
 1. How they figured out what the information says; and
 2. Whether they were interested in the information;
3. How CRs and community members understood the relationship of the information to a problem (quality of the agenda setting):
 1. How they picked a problem (including the degree of reliance on the facilitator);
 2. The degree to which the problem is related to health improvements;
 3. The breadth of the problem (whether it is a problem for much of the community, or only some parts of the community).

During the social action meeting, this includes:

1. Who (what demographic groups) attended meetings.
2. How CRs and community members understood the information presented about different kind of social actions that they might take (information and its salience):
 1. How they figured out what the information says; and

2. Whether they were interested in the information;
3. How CRs and community members figure out what they want to do with the information (quality of the planning):
 1. How CRs and community members understood the relationship of the information to the problem they determined in meeting 2
 2. The degree of local knowledge used in the planning;
 3. The degree of reliance on the facilitator and the examples of action furnished in the intervention; and
 4. The degree to which the discussion was broadly representative or dominated by a few people and excluded important points of view.
4. Planned actions to be undertaken by community members, individually or collectively, long or short route or combination, that have a plausible connection to the intervention:
 1. Individual actions;
 2. Group actions, and the degree of agreement among the group;
 3. The degree to which these actions are collaborative or oppositional;
 4. The degree to which these actions reflect the examples used in the intervention; and
 5. The relationship of group actions to existing or newly formed formal or informal organizations.

During the follow up meeting, this includes:

1. Who (what demographic groups) attended meetings.
2. Actions undertaken by community members, individually or collectively, long or short route or combination, that have a plausible connection to the intervention:
 1. Individual actions;
 2. Group actions, and the degree of agreement among the group;
 3. The degree to which these actions are collaborative or oppositional;
 4. The degree to which these actions reflect the examples used in the intervention;
 5. The relationship of group actions to existing or newly formed formal or informal organizations; and
 6. The interaction of contextual factors with these actions;
3. Provider or official sensitivity to these actions:
 1. The degree to which the targets of actions—e.g. providers or officials—ignore those actions or respond to them in some way; and
 2. The interaction of contextual factors with provider or official sensitivity;
4. The constructiveness of the provider or official response:
 1. Perceptions of the intent of the response (constructive or subversive);
 2. The interaction of contextual factors with the constructiveness of the response; and
 3. The degree to which the response fixed the problem;
5. Any evolution of community actions:
 1. The [revealed] accuracy of the community's expectations about what social actions would fix the problem

2. The degree to which the community (or civil society connected to the community) adapted—changed its approach—when its expectations about what social actions would be effective turned out not to be accurate.
3. The interaction of any adaptation in approach with the degree of empowerment the community perceives.
6. Actions taken around the aspects of health care dealt with in the intervention that have no plausible relationship to the intervention:
 1. Actions taken by community members to try to improve health care that are not connected to the intervention.

Key Informant Interviews (Layer 2)

Key informant interviews will be conducted with approximately 10 informants per village. The main purpose is to verify and provide more details on the social actions designed and undertaken by the community representatives.

Mechanisms, emergent and predicted (Question 5). Key informants will answer questions about the social actions, such as their involvement in the actions, opinion of the actions, and the current status of the actions. These interviews strive to assess:

1. Actions undertaken by community members, individually or collectively, long or short route or combination, that have a plausible connection to the intervention:
 1. The degree to which these actions are collaborative or oppositional;
 2. The relationship of group actions to existing or newly formed formal or informal organizations; and
 3. The interaction of contextual factors with these actions;
2. Provider or official sensitivity to these actions:
 1. The degree to which the targets of actions—e.g. providers or officials—ignore those actions or respond to them in some way; and
 2. The interaction of contextual factors with provider or official sensitivity;
3. The constructiveness of the provider or official response:
 1. Perceptions of the intent of the response (constructive or subversive);
 2. The interaction of contextual factors with the constructiveness of the response; and
 3. The degree to which the response fixed the problem;
4. Any evolution of community actions:
 1. The [revealed] accuracy of the community’s expectations about what social actions would fix the problem
 2. The degree to which the community (or civil society connected to the community) adapted—changed its approach—when its expectations about what social actions would be effective turned out not to be accurate.
 3. The interaction of any adaptation in approach with the degree of empowerment the community perceives.
5. Actions taken around the aspects of health care dealt with in the intervention that have no plausible relationship to the intervention:

1. Actions taken by community members to try to improve health care that are not connected to the intervention.

Context (Question 6). Key informants will answer select questions about context, such as their perception of how much of a problem health care is in the community.

In general, the Key Informant Interviews will also be an important source of data for process tracing the intervention.

Participation Monitoring (Layer 3)

Participation monitoring (PM) will take the form of a simple seating chart utilized during the scorecard meeting, social action meeting, and 90-day follow up meeting. Its purpose is to determine the nature and degree of participation in the intervention, and the data will be combined with empowerment survey data to determine if participation is related to any changes in CRs' perceptions of their empowerment.

Empowerment, power and efficacy (Question 4). Coders will record the distribution of influence and action across the intervention.

During all meetings this includes:

1. Participant arrival and exit;
2. Who talks;
3. Where participants sit;
4. Who acts aggressively;
5. Who volunteers;
6. Who is assigned activities
7. Participant movement throughout the meeting; and
8. Overall engagement.

During the follow up meeting, this includes:

1. Who reports having done an activity related to the intervention.

See Appendix B for an example seating chart.

Empowerment Survey (Layer 3)

The empowerment survey (ES) is administered to all CRs in the same communities that are observed using the PM tool. The ES is a short survey that combines an empowerment question with three anchoring vignettes, used to ascertain a baseline level of CR empowerment and later determine whether empowerment changed over the course of the intervention.

Empowerment, power and efficacy (Question 4)

Survey Module:

- (1) Do you feel that you have the power to make important decisions and take actions that improve life in this village, for yourself and others?

Rate yourself on a 1 to 4 scale, where 1 means being totally unable to improve life in this village, and 4 means having a lot of ability to improve live in this village.

- (2) Now I'm going to tell you about a situation in a village like yours. In this village, access to clean water is difficult. Water in the village is dirty and sometimes makes people sick. To get clean water, men and women have to walk 2 hours to a nearby village.

Now I'm going to describe the situation of three different people who want to improve this situation by digging a new well in the village. For each of them, I would like you to answer the question "how much ability does this person have to improve life in this village? Rate each person on a 1 to 4 scale, where 1 means being totally unable to improve life in this village, and 4 means having a lot of ability to improve life in this village.

A. Sri approached several of her neighbors to get support for her idea of digging a new well so that the village could have clean water. She and the neighbors then went to the village head's office and met with the village head. Sri was able to convince the village head to allocate some village funds for this purpose and also to make an announcement at some community meetings about the well project. After the announcement, many community members volunteered to help with the well, and Sri was able to buy materials for digging the well with the village funds. Together they were able to build the new well in just a few weeks.

B. Tati approached several of her neighbors to get support for her idea of digging a new well so that the village could have clean water. She and the neighbors then went to the village head's office and met the village head. At the meeting they found that the village fund was not available to help with the well. They made announcements about the well at several community meetings, but they were not able to get any volunteers because everyone was too busy. Tati did not give up, and after a lot of hard work she was able to collect money to buy the materials by asking for donations from several people in the village, and she was also able to convince some friends and neighbors to help work on the well themselves. After many months of working, they were finally able to complete the well.

C. Murni approached several of her neighbors to get support for her idea of digging a new well so that the village could have clean water. She and the neighbors then went to the village head's office and met the village head. At the meeting they found that the village fund was not available to help with the well. They

made announcements about the well at several community meetings, but they were not able to get any volunteers because everyone was too busy. Murni was disappointed and decided not to pursue her idea anymore. Instead she and her family continue to walk 2 hours to get clean water from a nearby village.

- (3) Now I'd like you to think again about your own ability to make important decisions and take actions that improve life in this village, for yourself and others. Which of the three people I just described is the most similar to your own ability to improve life in this village?

Ethnographic Data Collection (Layers 4).

An expanded set of descriptive elements will be gathered in the 6 communities with ethnographic researchers. These categories are similar to those in the RCT surveys and standard coding scheme; each category, however, includes additional elements, as the embedded ethnographic approach allows far more nuance and depth in understanding community context and its interaction with the intervention in each of the categories.

The inductive nature of the ethnographic approach in these communities requires some flexibility in the particular aspects. As described above, prior to entering the field, the ethnographic researchers will meet for several days as a group to jointly develop specific questions in each of the following categories. These preliminary questions will serve to focus the data collection efforts within each village and to organize the data collection across villages to facilitate comparison. Individual researchers will also be encouraged to explore additional questions and hypotheses that they develop in the field.

Empowerment, power and efficacy (Question 4). Ethnographers will seek to understand why community members, providers, and officials act the way they act and the relationships of power and efficacy reflected in those actions. This will involve understanding the local history and the evolution of the context before, during, and after the intervention:

1. Beliefs and practices of community members;
2. Beliefs and practices of providers and officials;
3. Perceptions of power and efficacy (subjective perceptions):
 - a. Which groups and individuals feel that their voice is taken seriously; and
 - b. The degree to which perceptions of efficacy influences groups' willingness or reluctance to engage in actions;
4. Distribution of influence and action across the intervention:
 - a. Who talks;
 - b. Where participants sit;
 - c. Whose problems gets raised;
 - d. Who is blamed for the problems;
 - e. Who takes responsibility for the response; and
 - f. The interaction of this distribution with contextual factors

Mechanisms, emergent and predicted (Question 5). Ethnographers will seek to understand what mechanisms, if any, the intervention triggered. During and after the intervention, ethnographers will seek to understand:

1. Participation in the prescribed actions of the intervention, including who attended meetings and why.
2. How CRs and community members understood the information (information and its salience):
 1. How they figured out what the information says;
 2. Whether they were interested in the information; and
 3. Whether this information fed out into the community or stayed only among participants;
3. How CRs and community members understood the relationship of the information to a problem (quality of the agenda setting):
 1. How they picked a problem (including the degree of reliance on the facilitator);
 2. The degree to which the problem is related to health improvements;
 3. The breadth of the problem (whether it is a problem for much of the community, or only some parts of the community); and
 4. Any differences in perceptions of the problem between participants and the broader community;
4. How CRs and community members figure out what they want to do with the information (quality of the planning):
 1. The degree of local knowledge used in the planning;
 2. The degree of reliance on the facilitator and the examples of action furnished in the intervention; and
 3. The degree to which the discussion was broadly representative or dominated by a few people and excluded important points of view.
5. Actions undertaken by community members, individually or collectively, long or short route or combination, that have a plausible connection to the intervention:
 1. Individual actions;
 2. Group actions, and the degree of agreement among the group;
 3. The degree to which these actions are collaborative or oppositional;
 4. The degree to which these actions reflect the examples used in the intervention;
 5. The relationship of group actions to existing or newly formed formal or informal organizations; and
 6. The interaction of contextual factors with these actions;
6. Provider or official sensitivity to these actions:
 1. The degree to which the targets of actions—e.g. providers or officials—ignore those actions or respond to them in some way; and
 2. The interaction of contextual factors with provider or official sensitivity;
7. The constructiveness of the provider or official response:
 1. Perceptions of the intent of the response (constructive or subversive);
 2. The interaction of contextual factors with the constructiveness of the response; and
 3. The degree to which the response fixed the problem;

8. The evolution of community actions:
 1. The accuracy of the community's expectations about what social actions would fix the problem
 2. The degree to which the community (or civil society connected to the community) adapted—changed its approach—when its expectations about what social actions would be effective turned out not to be accurate.
9. Actions taken around the aspects of health care dealt with in the intervention that have no plausible relationship to the intervention:
 1. Actions taken by community members to try to improve health care that are not connected to the intervention; and
 2. Actions taken by providers or officials to try to improve health care that are not connected to the intervention.

Each of these elements of the mechanism will be assessed across different contextual factors:

Context (Question 6). Ethnographers will seek to understand a number of elements of the context prior to, during, and after the intervention. These elements of context include:

1. Community characteristics:
 - a. Population
 - b. Education
 - c. Economy
 - d. Resource inequality
 - e. Gender
 - f. Ethnic ties
2. Health care characteristics:
 - a. Health conditions
 - b. Health services
 - c. Usage of health services
 - d. Cultural aspects of health, particularly around pregnancy and birth
 - e. Prices for services
 - f. Options for services across providers, including traditional providers
 - g. The perceived quality of health services
 - h. The perceived responsiveness of the system to community needs
 - i. The level of understanding within the community of each other's preferences and needs around health care.
3. Civil society:
 - a. Groups and Networks, including:
 - i. Informal associations
 - ii. Religious associations
 - iii. Local Non-Governmental Organizations
 - iv. International Non-Governmental Organizations
 - b. Membership and leadership of groups
 - c. Interactions of groups inside and outside the community
4. State institutions:

- a. Accountability structures and their degree of effectiveness at the community, regional, and national level;
 - b. Perceptions of the degree to which individual providers, officials, and politicians are responsive to community concerns; and
 - c. Political competitiveness, and the interaction of political actors with civil society actors and societal cleavages.
5. Configurations of local and traditional power:
- a. Perceptions of who typically takes responsibility for actions to address community problems; and
 - b. Perceptions of who typically holds accountable those responsible for making improvements.
6. Capacity for action, individual and/or collective:
- a. Participation in formal political activities
 - b. Participation in informal political activities (e.g. petitions, protests)
 - c. Social structures that facilitate individual and/or collective action
 - d. Perceptions of the efficacy of political action
7. Trust and solidarity
- a. Ability to borrow resources or money from others
 - b. Perceptions of those in the community
 - c. Trust in government officials
 - d. Willingness to contribute time or money to community projects that don't directly benefit the given individual
8. Social cohesion and inclusion
- a. Degrees of difference in the community
 - b. Differences that cause the most problems
 - c. Perceptions of safety
 - d. Dispute resolution mechanisms

Facilitator Reports (Layer 1—All Treatment Communities)

Information gathered by facilitators in the course of implementing the intervention will include some of the same information collected in the other layers; we will use this as supplementary data on community context and the implementation of the intervention:

Mechanisms, emergent and predicted (Question 5)

During the score card meeting, this includes:

- 1. Who (what demographic groups) attended meetings.

During the social action meeting, this includes:

- 1. Who (what demographic groups) attended meetings.
- 2. How CRs and community members figure out what they want to do with the information (quality of the planning):
 - 1. The degree to which the discussion was broadly representative or dominated by a few people and excluded important points of view.

3. Planned actions to be undertaken by community members, individually or collectively, long or short route or combination, that have a plausible connection to the intervention:

1. Copies of the social action plans;
2. The degree to which these actions are collaborative or oppositional.

During the follow up meeting, this includes:

1. Who (what demographic groups) attended meetings.
2. Planned actions to be undertaken by community members, individually or collectively, long or short route or combination, that have a plausible connection to the intervention:
 1. Copies of the revised social action plans (including new actions);
 2. The degree to which these actions are collaborative or oppositional;
 3. Challenges faced in the implementation of the actions.
3. How CRs and community members figure out how to address challenges (quality of the planning):
 1. The degree to which the discussion was broadly representative or dominated by a few people and excluded important points of view.

VII – Conclusion

Can information and discussion trigger citizen action to improve health outcomes? If so, how, and in what contexts? And does this process empower citizens? The Transparency for Development project is designed to answer these questions. It will seek to do so through a mixed-method evaluation, using a range of techniques from across the social sciences.

The organizing methodology is the randomized controlled trial. By implementing our intervention in a randomly selected set of communities and comparing them to a randomly selected set of control communities, we will be able to reliably estimate the effect of the intervention on health care utilization, the content of care, outcomes in maternal and newborn health, and perceptions of citizen empowerment and efficacy. RCTs of 200 communities in Tanzania and Indonesia will allow such estimates in two very different contexts.

To provide an understanding of the mechanisms behind any effects, as well as the role of context in conditioning those mechanisms, we will treat the 400 communities—200 in each country—as 400 cases, each with its own history, culture, demographic and economic characteristics, patterns of social interaction, and relationship with state institutions. We will study these cases using multiple additional methods—direct observation, focus groups, informant interviews, systematic coding of meetings, and ethnographic methods—with the goal of identifying patterns: for example, in the types of actions that certain communities choose to pursue, or the types of actions that seem to lead to greater improvements in health. None of these techniques will provide perfectly reliable information on communities or the processes of the intervention; nor will they allow findings about which we can be as confident as the findings of the randomized controlled trial. But by integrating them with each other and with the RCT, we hope to generate a wealth of additional overlapping and mutually-reinforcing information that will permit us to understand elements of the communities and processes of the intervention that are difficult to understand with the RCT. This includes our initial hypotheses about how our intervention will play out in communities with different contextual features (the “five worlds” framework). But the design also allows for emergent categories and processes from the field that we could never have anticipated *ex ante*. We hope that this combination of deductive and inductive techniques will allow us not only to check against the biases and imperfections of any particular methodology in testing our hypotheses, but also to check our own biases and assumptions in designing this project.

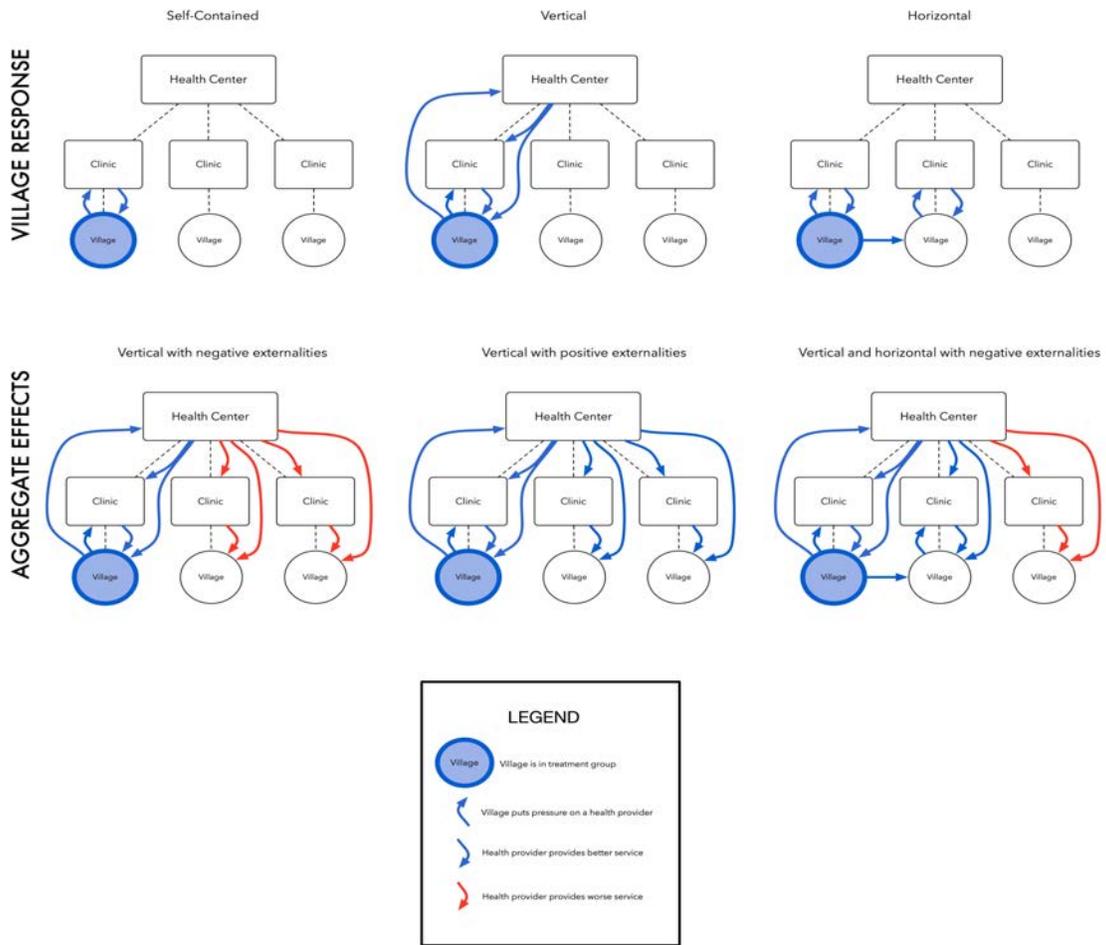
The primary aim of this mixed-methods design is a set of rigorous, reliable estimates of the impact of our intervention and an understanding of the processes and role of context in that impact. But in addition, this combination of methods should allow us to conclude this phase of the project with a set of hypotheses about types of communities in which information generates social action—both in general and of specific kinds—as well as the types of social actions that are consistently associated with greater improvements in health. These hypotheses can in turn form the basis for further exploration and evaluation. Likewise, we hope that the mixed-methods approach—particularly our

efforts to understand and analyze the essential types underlying community context—will permit the development of simple, straightforward tools for NGOs and practitioners to more easily identify the types of communities they are working in. And finally, we hope that the case studies themselves will prove useful, both for those wishing to understand the experience of a wide variety of communities experiencing an informational and community-action intervention, and those wishing to check the conclusions and interpretations we ultimately draw. The overriding goal is that this evaluation generates a wide range of reliable and useful evidence for practitioners, researchers, and stakeholders working to improve health, accountability, and citizen participation.

References

- Anderson, Michael L. 2008. "Multiple Inference and Gender Differences in the Effects of Early Intervention: A Reevaluation of the Abecedaian, Perry Preschool, and Early Training Projects." *Journal of the American Statistical Association* no. 103 (484):1481-1495.
- Casey, Katherine, Rachel Glennerster, and Edward Miguel. 2012. "Reshaping Institutions: Evidence on Aid Impacts Using a Preanalysis Plan." *Quarterly Journal of Economics* no. 127 (4):1755-1812.
- De Renzio, P. 2006. *Mutual Accountability*. ODI Briefing Paper, ODI, London.
- Eyben, Rosalind. 2010. "Hiding Relations: The Irony of 'Effective Aid'." *European Journal of Development Research* no. 22 (3):382-397. doi: 10.1057/ejdr.2010.10.
- Fung, Archon, Mary Graham, and David Weil. 2007. *Full Disclosure: The Perils and Promise of Transparency*. New York: Cambridge University Press.
- Grootaert, Christiaan, Deepa Narayan, Veronica Nyhan Jones, and Michael Woolcock. 2004. *Measuring Social Capital: An Intergrated Questionnaire*. Working Paper No. 18, World Bank.
- Joshi, Anuradha, and Peter P Houtzager. 2012. "Widgets or Watchdogs?" *Public Management Review* no. 14 (2):145-162. doi: 10.1080/14719037.2012.657837.
- King, Gary, Robert O. Keohane, and Sidney Verba. 1994. *Designing social inquiry: scientific inference in qualitative research*. Princeton, N.J.: Princeton University Press.
- Kling, Jeffrey R., Jeffrey B. Liebman, and Lawrence F. Katz. 2007. "Experimental Analysis of Neighborhood Effects." *Econometrica* no. 75 (1):83-119.
- Kosack, Stephen, and Archon Fung. 2014. "Does Transparency Improve Governance?" *Annual Review of Political Science*.
- Mercer, C. 2002. "NGOs, civil society and democratization: A critical review of the literature." *Progress in Development Studies* no. 2:5-22.
- Pagel, C, A Prost, S Lewycka, S Das, T Colbourn, R Mahapatra, K Azad, A Costello, and D Osrin. 2011. "Intracluster correlation coefficients and coefficients of variation for perinatal outcomes from five cluster-randomised controlled trials in low and middle-income countries: results and methodological implications." *Trials* no. 12 (51).
- Ranson, S. 2003. "Public accountability in the age of neo-liberal governance." *Journal of Education Policy* no. 18 (5):459-480.
- Scott, James C. 1987. *Weapons of the Weak: Everyday Forms of Peasant Resistance*. New Haven: Yale University Press.
- Scott, James C. 1999. *Seeing Like a State: How Certain Schemes to Improve the Human Condition Have Failed*. New Haven: Yale University Press.

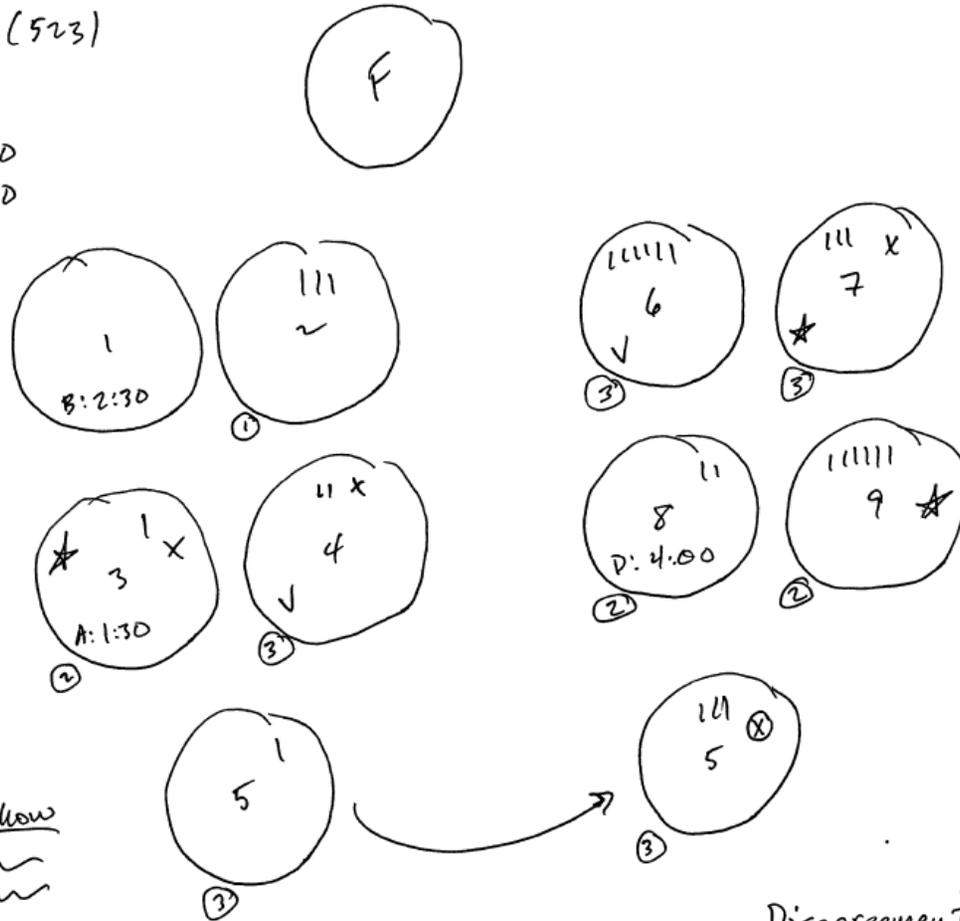
Appendix A – Spillover Design



Graphic design credit: Theodore Svoronos

Appendix B – Participation Monitoring Seating Chart Example

Date: 5/2/15
 Village: Uuu (523)
 District: Uuu
 Facilitator: Uuu
 Start time: 12:30
 End time: 5:30

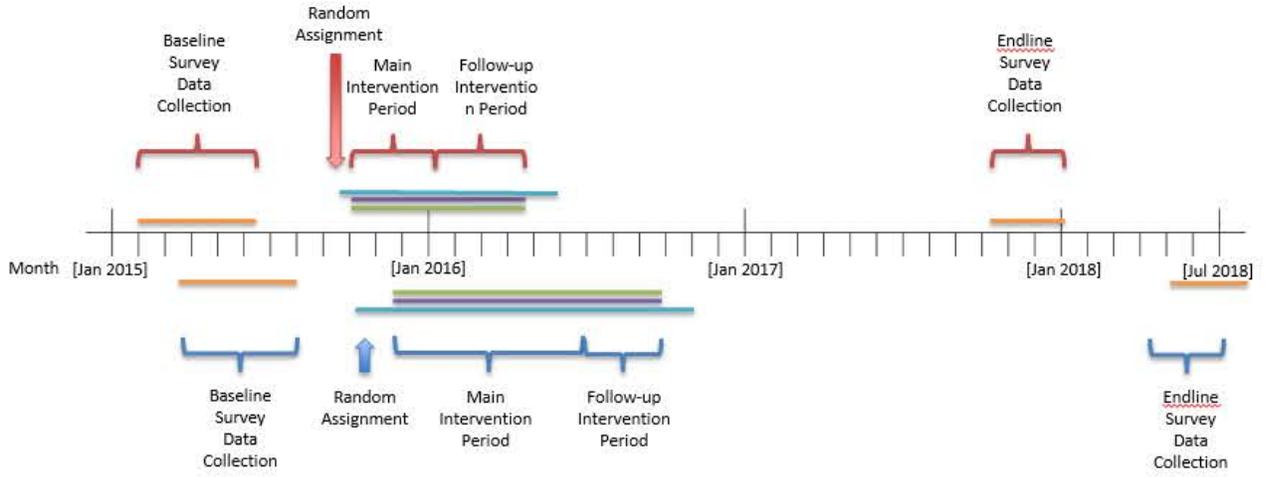


Participants

- | | | |
|---|-----|-----|
| 1 | Uuu | (M) |
| 2 | Uuu | (M) |
| 3 | Uuu | (F) |
| 4 | Uuu | (F) |
| 5 | Uuu | (F) |
| 6 | Uuu | (F) |
| 7 | Uuu | (F) |
| 8 | Uuu | (M) |
| 9 | Uuu | (M) |
- No show
- | | |
|---|-----|
| 1 | Uuu |
| 2 | Uuu |
| 3 | Uuu |
| 4 | Uuu |

Disagreement:
 3/4 vs. 7

Appendix C – Timeline



KEY	
■	Indonesia
■	Tanzania
■	Layer 4: Ethnography
■	Layer 3: PM, ES
■	Layer 2: SCS, KII
■	Layer 1: Baseline and Endline Surveys