Power to the People:

Evidence from a Randomized Field Experiment on Community-Based Monitoring in Uganda

This version: January 2008

Martina Björkman* and Jakob Svensson#

Abstract: This paper presents a randomized field experiment on community-based monitoring of public primary health care providers in Uganda. Through two rounds of community meetings, local NGOs encouraged communities to be more involved with the state of health service provision and strengthened their capacity to hold their local health providers accountable for performance. A year after the intervention, treatment communities are more involved in monitoring the provider and health facility staff exert higher effort to serve the community. We document large increases in utilization and improved health outcomes (reduced child mortality and increased child weight) that compare favorably to some of the more successful community-based intervention trials reported in the medical literature.

This project is a collaborative exercise involving many people. Foremost, we are deeply indebted to Frances Nsonzi and Ritva Reinikka for their contributions at all stages of the project. We would also like to acknowledge the important contributions of Gibwa Kajubi, Abel Ojoo, Anthony Wasswa, James Kanyesigye, Carolyn Winter, Ivo Njosa, Omiat Omongin, Mary Bitekerezo, and the field and data staff with whom we have worked over the years. We thank the Uganda Ministry of Health, Planning Division, the World Bank's Country Office in Uganda, and the Social Development Department, the World Bank, for their cooperation. We are grateful for comments and suggestions by Paul Gertler, Esther Duflo, Abhijit Banerjee, and seminar and conference participants at LSE, Oxford, IGIER, MIT, World Bank, NTNU, Namur, CEPR/EUDN conference in Paris, and BREAD & CESifo conference in Venice. Finally, we wish to thank the Bank-Netherlands Partnership Program (BNPP), the World Bank Research Committee, the World Bank Africa Region division, and the Swedish International Development Agency, Department for Research Cooperation, for funding this research. Björkman also thanks Jan Wallander's and Tom Hedelius' Research Foundation for funding.

*IGIER, University of Bocconi, and CEPR. Email: mbjorkman@unibocconi.it.

#IIES, Stockholm University, NHH, and CEPR. Email: jakob.svensson@iies.su.se.

1 Introduction

Approximately 11 million children under five die each year. Almost half of these deaths occur in sub-Saharan Africa where roughly one in five children dies before reaching the age of five. More than half of these children – nearly 6 million – will die of diseases that could easily have been prevented or treated if the children had had access to a small set of proven, inexpensive services.[1]

Why are these services not provided? While there is no simple answer, a wealth of anecdotal, and recently more systematic, evidence shows that the provision of public services to poor people in developing countries is constrained by weak incentives of service providers – schools and health clinics are not open when supposed to; teachers and health workers are frequently absent from schools and clinics and, when present, spend a significant amount of time not serving the intended beneficiaries; equipment, even when fully functioning, is not used; drugs and vaccines are misused; and public funds are expropriated.[2]

The traditional approach to accountability in the public sector relies on external control. This is a top-down approach where someone in the institutional hierarchy is assigned to monitor, control and reward/punish agents further down in the hierarchy. The tacit assumption is that more and better enforcement of rules and regulations will strengthen providers' incentives to increase both the quantity and quality of service provision. But, in many poor countries, the institutions assigned to monitor the providers are typically weak and malfunctioning, and may themselves act under an incentive system providing little incentives to effectively monitor the providers. As a result, the relationship of accountability of provider-to-state is ineffective in many developing countries.

As a complementary strategy, it has therefore been argued that more effort must be placed on strengthening beneficiary control, i.e. strengthening providers' accountability to citizen-clients (see e.g., World Bank, 2003). However, despite the enthusiasm for such an approach, there is little credible evidence on the impact of policy interventions aimed at achieving it (Banerjee and He, 2003; Banerjee and Duflo, 2005). This paper attempts to provide some.

[1] It is estimated that 2 million children under five die from diarrhea, which in most cases can be treated with simple oral rehydration therapy. Another 2 million children die from pneumonia, where once more there is sufficient evidence of effective treatment (antibiotics). Malaria kills one million children under five, most of whom could have been protected by preventive measures and treatment with antimalarials. Globally, neonatal disorders account for the highest proportion of deaths of children – many of them could have been saved if mothers had had access to basic antenatal and delivery care. Approximately half a million children under five die from measles, for which these is a cheap and effective vaccine (Black et al., 2003; Jones et al., 2003).

[2] For anecdotal and case study evidence, see World Bank (2003). Chaudhury et al. (2006) provide systematic evidence on the rates of absenteeism based on surveys where enumerators made unannounced visits to primary schools and health clinics in seven developing countries. Averaging across countries, 35 percent of the health workers were absent. Banerjee et al. (2004) and Duflo and Hanna (2005) confirm these findings. On misappropriation of public funds and drugs, see Reinikka and Svensson (2004) and McPake et al. (1999).

To examine whether beneficiary control works, we designed and conducted a randomized field experiment in 50 "communities" from nine districts in Uganda.[3] In the experiment, small local NGOs facilitated two rounds of village and staff meetings. During the meetings, the community discussed and analyzed baseline information on the status of health service delivery relative to other providers and the government standard. Community member were also encouraged to develop a plan identifying key problems, priorities, and possible solutions. The primary objective of the intervention was to kick-start a process (of community monitoring) which was then up to the community to sustain and lead.

The community-based monitoring project increased the quality and quantity of primary health care provision. A year after the first round of meetings, we find a significant difference in the weight of infants (0.17 z-score increase) and a markedly lower number of deaths among children under five (33 percent reduction in child deaths) in the treatment communities. Utilization (for general outpatient services) was 16 percent higher in the treatment compared to the control facilities and the overall effect across a set of utilization measures (deliveries, use of antenatal care, and family planning) is large and significantly positive. Treatment practices, as expressed both in perception-based responses by households and in more quantitative indicators (immunization of children, waiting time, examination procedures, absenteeism), improved significantly in the treatment communities, thus suggesting that the changes in quality and quantity of health care provision are due to behavioral changes of the staff. We find evidence that the treatment communities became more engaged and began to monitor the health unit more extensively. No effect is found on investments, level of financial, or in-kind support from the government. Furthermore, supervision of providers by upper-level government authorities remained low in both the treatment and the control group. This reinforces our confidence that the findings on the quality and quantity of health care provision resulted from increased efforts by the health unit staff to serve the community in the light of better community monitoring.

Although research on medical interventions (i.e. biological agents such as vaccines and drugs, or treatment practices) is plentiful, little is known about the characteristics of delivery strategies capable of achieving and maintaining high coverage for specific interventions, or packages of interventions, in various epidemiological, health system, and cultural context (Bryce, et al., 2003). In this paper, we focus on one mechanism that have been highlighted, but not examined, in the literature - - a mechanism of accountability enabling (poor) people to scrutinize whether or not those in authority have fulfilled their health responsibilities.[4]

This paper also relates to a small literature on improving governance and public service delivery through community participation. Olken (2005) evaluates different ways of monitoring corruption in a road construction project in Indonesia. In one of the experiments, invitations were sent out to village-level meetings where project

[3] A "community" is operationalized as the households (and villages) residing in the five-kilometer radius around the facility (see section 5 for details). Approximately 110,000 households (600,000 individuals) reside in these communities, of which half reside in the treatment communities.

[4] See, for example, Yamey et al. (2007).

officials documented how they spent funds for local road construction. He only finds minor effects of the intervention. Our work differs in important dimensions. First, unlike Olken's study, where the meetings were typically dominated by members of the village elite, in our project the meetings facilitated by the local NGOs were organized and structure in a way to avoid elite capture. Second, unlike corruption which is not easily observable[5], the information discussed in the meetings were basic facts, in absolute and relative terms, on utilization and quality of services based on the community's own experience. Finally, the intervention we evaluate sought to address two constraints highlighted in the literature on community participation/monitoring: lack of relevant information and inadequate participation. Using a randomized design, Banerjee, Deaton and Duflo (2004) evaluate a project in Rajasthan in India where a member of the community was paid to check once a week whether the auxiliary nurse-midwife assigned to the health center was present at the center. The intervention had no impact on attendance and the authors speculate that a key reason for this is that the individual community member, although informed, did not manage to use his or her information on absenteeism to invoke community participation. Here, on the contrary, we explicitly try to address this participation constraint, by involving a significant number of community members and encourage them to jointly develop a monitoring plan.

Finally, the paper also links to a growing empirical literature on the relationship between information dissemination and accountability. With few exceptions, this literature studies the relationships of accountability of politicians to citizens and deal with one (periodic elections) mechanism through which citizens can make politicians and policymakers accountable (see for instance Strömberg, 2004; Besley and Burgess, 2002; Ferraz and Finan, 2005). Our work differs in several dimensions. First, we focus on mechanisms through which citizens can make providers, rather than politicians, accountable. Thus, we do not study the design or allocation of public resources across communities or programs, but rather on how these resources are utilized. Second, we use micro data from households and health stations rather than disaggregated national accounts data. Finally, we identify impact using an experimental design.

The next section briefly describes the institutional environment in Uganda and in the project areas. The community-based monitoring intervention is described in section 3. Section 4 lays out the evaluation design and the results are presented in section 5. Section 6 concludes.

2 Institutional Setting

Uganda, like many newly independent countries in Africa, had a functioning health care system in the early 1960s. Accessibility and affordability were relatively extensive. The 1970s and 1980s saw the collapse of Government services as the country underwent

[5] Olken also reports that corruption problems were seldom discussed in these meetings.

political upheaval. Health indicators fell dramatically during this period until peace was restored in the late 1980s. Since then, the Government has been implementing major infrastructure rehabilitation programs in the public health sector. Some health indicators have improved, while others have not. This is despite a GDP growth rate exceeding 64 percent and a 40-percent reduction in consumption poverty in the 1990s (Appleton 2001)

As of 2001, public health services are free of charge. Anecdotal and survey evidence (see below), however, suggests that users still encounter varying costs when visiting public health facilities.

The health sector in Uganda is composed of four types of facilities: hospitals, health centers, dispensaries (health center III), and aid posts or sub-dispensaries. These facilities can be government, private for-profit, or private not-for-profit operated and owned. The impact evaluation focuses on dispensaries. Dispensaries are in the lowest tier of the health system where a professional interaction between users and providers takes place. Most dispensaries are rural (89 percent). According to the government health sector strategic plan, the standard for dispensaries includes preventive, promotional, outpatient care, maternity, general ward, and laboratory services (Republic of Uganda 2000). In our sample of facilities, on average, a dispensary was staffed by an in-charge or clinical officer (a trained medical worker), two nurses (including midwives), and three nursing aids or other assistants.

The health sector in Uganda is decentralized and supervision and control of the dispensaries are governed at the district level. A number of actors are responsible for the functioning of the dispensaries. The Health Unit Management Committee (HUMC) is supposed to be the main link between the community and the health facility. Each dispensary has an HUMC, which consists of members from both the health facility staff and non-political representatives from the community (elected by the sub-county local council). The HUMC should monitor drugs and finances disbursed to the health facility, as well as the day-to-day running of the health facility (Republic of Uganda 2000). The HUMC can warn the health facility staff on issues of indiscipline, rudeness to patients and misappropriations of funds and recommend that a worker is transferred from the health facility. However, the HUMC has no authority to dismiss a worker. In cases of problems at the health facility, the working practice is that the chairperson of the HUMC raises the issue with the in-charge. If there is no improvement, the issue should be referred to the Health Sub-district.

The Health Sub-district monitors funds, drugs and service delivery at the dispensary. Supervision meetings by the Health Sub-district are supposed to appear quarterly but, in practice, monitoring is infrequent. The Health Sub-district has the authority to reprimand, but not dismiss, health facility staff for indiscipline. In severe cases of indiscipline, therefore, the errand will be referred to the Chief Administrative Officer of the District and the District Service Commission, which is the appointing authority for the district and has the authority to suspend or dismiss staff.

Various local NGOs, so-called Community-based organizations (CBOs), are also active in the primary health care sector. These CBOs mainly focus on health education

in antenatal care, family planning, and HIV/AIDS prevention.

3 The Project: Citizen Report Card

3.1 Overview and expected results of the intervention

In response to perceived continued weak health care delivery at the primary level, a pilot project (Citizen report cards) aimed at enhancing community involvement and monitoring in the delivery of primary health care was initiated in 2004. The project was designed by staff from Stockholm University and the World Bank, and implemented in cooperation with a number of Ugandan practitioners and 18 community-based organizations, or local NGOs (we use the two terms interchangeably). The 50 project facilities (all in rural areas) were drawn from nine districts in Uganda (see the working paper version for details).

With the catchment area (or the community) of each dispensary defined as the households and villages residing within a five-kilometer radius from the clinic, about 110,000 households reside in the communities supposedly served.[6] The facilities were first stratified by location (districts) and then by size (the number of households residing in the catchment areas). From each group, half the units, with corresponding catchment areas, were randomly assigned to the treatment group and the remaining 25 units were assigned to the control group. Each district thus had both treatment and control groups.

The main objective of the Citizen report card project was to strengthen providers' accountability to citizen-clients by enhancing communities' ability to monitor providers. The intention was to initiate a process, using trained local actors (CBOs) as facilitators, which the communities themselves could manage and sustain. To this end, the project aimed at: (i) providing communities with information on relative performance; and (ii) encouraging people to develop a plan that identified steps the provider and the community should take to improve service performance and ways to get the community more actively involved in monitoring the provider. These components are discussed in the following sub-sections. A time-line and a schematic view of the intervention and expected outcomes are depicted in figures 1 and 2.

The key behavioral change induced by more extensive community-based monitoring was expected to be increased effort by health unit staff to serve the community. In Uganda, as in many other developing countries, health workers have little pecuniary incentives to exert high effort. Public money does not follow patients and hiring, salaries and promotions are largely determined by seniority and educational qualifications – not by how well the staff performs. While formal sanctions, such as suspensions and

[6] Dispensaries are designed to serve households in a catchment area roughly corresponding to the five-kilometer radius around the facility (Republic of Uganda, 2000).

dismissal, are possible, they are in practice uncommon and only applied in cases of severe neglect and mismanagement. An individual worker may of course still put in high effort if shirking deviates from her ideal choice, given the behavior of other staff and the situation (Akerlof and Kranton, 2005). The effort choice may also be influenced by social rewards from community members or social sanctions against shirking health workers. Social rewards and sanctions are key instruments available to the community to boost health worker's effort.

However, rural communities typically lack access to reliable and structured information on their entitlements and the status of service delivery to systematically use these instruments. Although people know whether their own child died or not, and whether the health workers did anything to help them, they typically do not have any information on aggregate outcomes, such as how many children in their community did not survive beyond the age of 5 or where citizens, on average, seek care (Khemani, 2006). Provision of information on outcomes and performance improves citizens' ability to challenge abuses of the system, since reliable quantitative information is more difficult for service providers to brush aside as anecdotal, partial, or simply irrelevant. Therefore, as the community receives more accurate information about service quality and can coordinate on expected reforms, i.e. the intervention, we expected the community (and individual members) to be in a better position to monitor effort and thereby choose to more systematically exploit the instruments at their disposal, i.e., praise workers when service provision improves and complain when it does not. Workers may then find coming to work, or more generally exerting effort, more attractive. As service quality improves, we anticipated community members in turn to shift from self-treatment to the facility in question. The switch from self-treatment to professional care and the increase in quality could both have a positive effect on health outcomes.

3.2 Data collection and report cards

Data collection was governed by two objectives. First, data were required to assemble report cards on how the community at large views the quality and efficacy of service delivery. We also wanted to contrast the citizens' view with that of the health unit staff. Second, data were required to rigorously evaluate impact. To meet these objectives, two surveys were implemented: a survey of health care providers and a survey of health care users. Both surveys were implemented prior to the intervention (data from these pre-intervention surveys formed the basis for the report cards) and one year after the project had been initiated.

A quantitative service delivery survey was used to collect data from the health service providers. Since agents in the service delivery system may have a strong incentive to misreport (or not report) key data, the data were obtained directly from the records kept by facilities for their own need (i.e. daily patient registers, stock cards, etc.) rather than from administrative records submitted to the district-level government. The former, often available in a highly disaggregate format, were considered to

suffer the least from any incentive problems in record-keeping.

The household survey collected data on both households' health outcomes and health facility performance, including performance parameters such as usage, availability, access, reliability, quality and satisfaction. To the extent that it was possible, household responses were supported by patient records (i.e., patient exercise books and immunization cards). These records helped the household recall details about its visits to the health facility and also minimized problems of misreporting. The post-intervention household survey also included a shorter module on health outcomes. Specifically, data on under-five mortality were collected and we measured the weight of all infants in the surveyed households.

A stratified random sample of households within the catchment area of the facility were surveyed. In total, roughly 5,000 households have been surveyed in each round. The design and implementation of the surveys are explained in more detail in the working paper version of this paper and summary statistics are reported in appendix.

The data from the two pre-intervention surveys were analyzed and a smaller subset of the findings were assembled in report cards for the treatment localities.[7] The data included in the report cards were identified as key areas subject to improvement and include utilization, quality of services, informal user charges and comparisons vis-à-vis other health facilities in the district and the country at large. Each treatment facility and its community had a unique report card summarizing, in a format easily accessible to the communities, the key findings from the surveys conducted in their area.

The report cards were translated into the main language spoken in the community.[8] To support the non-literate community members, posters were designed by a local artist so that otherwise complex information and concepts were easily understood. Because the information in the report cards was largely statistical, the posters visually conveyed the main messages, such as where people go to seek medical care and why they do so.

3.3 Dissemination and participation

Getting people to retain and use information to achieve a specific objective is a complex problem.[9] Extensive piloting concluded that simply reporting baseline information on service delivery outcomes would be likely to have little impact. Thus, to maximize the likelihood that the information in the report cards would be used when people decide what actions to take, a participatory approach was chosen where community members themselves actively interpreted and analyzed the information. To this end, the process

[7] Thus, the design and size of the surveys were largely driven by the second objective – to evaluate impact.

[8] In the end, the report cards were translated into six different languages: Ateso (Soroti), Lusoga (Iganga), Lango (Apac), Luganda (Masaka, Wakiso, Mukono and Mpigi), Runyankore (Mbarara) and Lugbara (Arua).

[9] See, for example, Lupia (2004) who systematizes and draws conclusions from clinical, psychological, and economic research on information transmission and processing.

of providing information and encouraging participation was initiated through a series of meetings: a community meeting; a staff meeting; and an interface meeting. Staff from various local NGOs (CBOs) acted as facilitators in these meetings.[10]

The community meeting was a two-day (afternoons) event with approximately 100 invited participants drawn from the surveyed villages in the catchment area of the health facility. To avoid elite capture, the invited participants consisted of a selection of representatives from different spectra of society (i.e. young, old, disabled, women, mothers, leaders). The facilitators mobilized the village members by cooperating with village council representatives in the catchment area. Invited participants were asked to spread the word about the meeting and, in the end, a large number of uninvited participants who had found out about the event also attended the meeting. A typical village meeting was attended by more than 150 participants per day.

In the community meeting, the facilitators used a variety of methods to disseminate the information in the report cards in a participatory, or interactive, way.[11] Information on patients' rights and entitlements was also discussed.[12] As the objective was not only to inform but to encourage people to participate in developing a shared view on how to improve service delivery and monitor the provider, the facilitators structured the discussions through a series of questions on the various elements of accountability in the primary health sector (*who* is accountable to *whom*?; *what* is a particular actor accountable *for*?; *how* can these actors account for their *actions*?; and how are these elements reflected in the report card findings?). During the discussion, the participants were divided into focus groups (women, men, older, leaders, and youths) so that also more marginalized groups such a women and youth could raise their voices and discuss issues specific to their group. At the end of the meeting, the community's suggestions for improvements (and how to reach them without additional resources) were summarized in an action plan. The action plan contained information on health issues/services that had been identified by the community as the most important to address; how these issues could be addressed and how the community could monitor improvements (or lack thereof). While the issues raised in the action plans differed across communities, a common set of concerns included high rates of absenteeism, long waiting-time, weak attention of health staff, and differential treatment. After the

[10] Since the CBOs are situated in these rural communities and had a mandate drawn from a long-term presence on the ground working with the community, these facilitators were perceived to be a good conduit through which the project could be delivered. The CBO facilitators were trained for seven days in data interpretation and dissemination, utilization of the participatory methodology, and conflict resolution and management. It should be noted that various CBOs (including some participating in the project) also operate in the control districts. Thus, the presence (and numbers) of CBOs in the project communities is similar across treatment and control groups.

[11] See the appendix for a more detailed description of the various methods used during the meetings.

[12] Information on patients' rights and entitlements was based on the Yellow Star program. In 2000, the MoH developed a quality of care strategy called the Yellow Star Program with the aim of improving and maintaining basic standards of care at government health facilities. The Yellow Star Program lists a set of basic standards of quality. The standards fall into six categories: Infrastructure and Equipment; Management systems; Infection prevention: Information: Education and Communication; Clinical skills; and Client services.

meeting, participants were given posters and copies of the report card to bring back to their villages and share with their village members.

The health facility staff meeting was a one-day (afternoon) meeting held at the health facility with all staff present. In this meeting, the facilitators contrasted the information on service provision as reported by the provider with the findings from the household survey. The meeting enabled the providers to review and analyze their performance, and compare their performance with other health clinics in the district and across the country.

An interface meeting with participants (chosen by people that attended the community meeting) from villages in the catchment area and the health facility staff followed the community and health facility meetings. During the interface meeting, the community representatives and the health facility staff presented and discussed their suggestions for improvements. A role-playing exercise was used to visualize (and defuse tensions) the current situation at the health facility, with community participants and staff reversing roles. The participants discussed their rights and responsibilities as patients or medical staff. The outcome was a shared action plan, or a contract, outlining the community's and the service provider's agreement on what needs to be done, how, when and by whom. The "community contract" also identified how the community could monitor the agreements and a time plan. Because the problems raised in the community meetings constituted the core issues discussed during the interface meetings, the community contract was in many respects similar to the community's action plan. Copies of the community contract were kept with the community and the health facility to support the following monitoring process.

3.4 Ongoing process of monitoring

The three separate meetings aimed at kick-starting the process of community monitoring. Thus, after the initial meetings the communities were themselves in-charge of establishing ways of monitoring the provider. As an integrated part of the CBO's ordinary work in the villages, the facilitators were asked to support, if possible, the community in this process (approximately two follow-up visits in the six-month period that followed). However, because there was no outside presence in the communities, we cannot verify if these support visits actually took place.

After a period of six months, the communities and health facilities were revisited to conduct a mid-term review – a repeat engagement on a smaller scale. Including a one-day community meeting and a one-day interface meeting facilitated by the local NGOs, the review tracked the implementation of the community contract. Health facility staff and community members jointly discussed suggestions on actions for sustaining or improving progress, or in the case of no improvements, why so. Where improvements had been made, suggestions for sustainability were recorded. The community and the health facility kept the updated community contract to assist in further monitoring.

4 Evaluation Design and Expected Outcomes

4.1 Outcomes

The main outcome of interest is whether the intervention increased the quantity and quality of health care and, thus, improved health outcomes in the treatment communities. However, we are also interested in evaluating changes (if any) in all steps in the accountability chain depicted in figure 2: Did the intervention increase treatment communities' ability to exercise accountability? Did it result in behavioral changes of the staff (i.e., did staff exert higher effort to serve the community)?

As a robustness test, we also assess alternative explanations. Some of these alternative mechanisms are illustrated in figure 3. One concern is spillovers. Spillovers could affect the estimates in two ways. If information about the intervention spread to control areas and, as a result, control communities became more involved in monitoring the providers, the estimated treatment effect would be biased downward. If, on the other hand, households in control communities shifted from seeking care at the control facility to the nearest treatment clinic, it is possible that the estimated treatment effect would be biased upward. This is a potentially serious concern but also a mechanism which we can test. It is also possible that the intervention did not only (or primarily) increase the extent of community monitoring, but had an impact on other agents in the service delivery chain. For example, the various upper-level authorities in the health sector (e.g. the Health Sub-district) may have become more involved in monitoring the providers, or the district government may have increased its administrative or financial support, following the intervention. While this would not invalidate the causal effect of the intervention it would, of course, affect interpretation. Therefore, this alternative hypothesis is also subject to a battery of tests.

Given the wealth of information, we report the main results and tables in the text and refer the reader to the working paper version and appendix for additional findings.

4.2 Statistical framework

Given the randomized assignment of the Citizen Report Card project, we expect the 2004 pre-data in the treatment areas to be similar those in the control areas. We have both facility-specific data (on utilization, for example) and household-specific data (on waiting time, for example). Denoting y_{ijdt} the outcome variable of household i (when applicable), health facility j in district d and period t, we start by checking that there is no difference between treatment and control facilities/communities prior to the intervention:

$$y_{ijdPRE} = \alpha_{PRE} + \beta_{PRE} T_{jd} + \varepsilon_{ijdPRE} \ , \tag{1}$$

where $t = PRE$ denotes the pre-intervention period, T_{jd} is a dummy indicating whether health facility j is in the treatment group and ε_{ijdPRE} is the error term.

In regressions using household data, the disturbance term is adjusted to allow for correlations within catchment areas (communities).

To estimate the causal effect of the program, we then run the same regression in the post-period ($t = POST$):

$$y_{ijdPOST} = \alpha_{POST} + \beta_{POST} T_{jd} + \varepsilon_{ijdPOST} \ . \tag{2}$$

We also estimate an extended version of equation (2):

$$y_{ijdPOST} = \alpha + \beta_{POST} T_{jd} + X_{ijdPOST}\pi + \theta_d + \varepsilon_{ijdPOST} \ . \tag{3}$$

Specification (3) includes district fixed effects (θ_d) and facility and household variables (X). Due to random assignment, T should be orthogonal to X, and the consistency of β_{POST} does not depend on the inclusion of X in the model. Regression adjustment in equation (3) is used to improve estimation precision and to account for stratification and chance differences between groups in the distribution of pre-random assignment (Kling et al., 2004).

For a subset of variables, we can also stack the pre and post data and explore the difference-in-differences in outcomes, i.e., we estimate:[13]

$$y_{ijt} = \gamma POST_t + \beta_{DD}(T_j * POST_t) + \mu_j + \varepsilon_{ijt}, \tag{4}$$

where $POST$ is a post period dummy, μ_j is a facility specific fixed effect, and β_{DD} is the difference-in-differences estimate (program impact).[14]

For some outcomes, like utilization, we have several measures (out-patients, deliveries, antenatal care visits, family planning visits). To form judgment about the impact of the intervention on a family of K related outcomes, we also calculate average standardized treatment effects, $\tilde{\beta} = \frac{1}{K}\sum_{k=1}^{K} \frac{\hat{\beta}_k}{\hat{\sigma}_k}$, where $\hat{\beta}_k$ and $\hat{\sigma}_k$ are the point estimate and standard error, respectively, for each effect (see Duflo et al., 2006). Specifically, we follow Kling et al. (2004) and estimate a seemingly unrelated regression system,

$$Y = [I_K \otimes (T \ X)]\,\theta + \upsilon, \tag{5}$$

where I_K is a K by K identity matrix. We calculate point estimate, standard error, and p-value for $\tilde{\beta}$ based on the parameters, $\hat{\beta}_k$ and $\hat{\sigma}_k$, jointly estimated as elements of θ in (5).

[13] It is a subset of variables since the post intervention surveys collected information on more variables and outcomes.

[14] A slightly more restricted difference-in-difference (DD) specification substitutes the facility fixed effects for T_{jd}. In that case, time invariant factors will be captured by T_{jd}. Both DD specifications yield identical point estimates of β_{DD}.

5 Results

5.1 Pre-intervention differences

Prior to the intervention, the treatment and the control group were similar on most characteristics. We report the test of difference in means across control and treatment groups in table 1. At baseline, we do not find any statistically significant differences in utilization (number of outpatient treated and deliveries per month), households' use of different service providers (including drug shops) in case of illness, waiting time, equipment usage, government funding of clinics, citizens' perceptions of staff behavior, catchment area characteristics (such as the number of villages and households in catchment area), distances from the health facility to the nearest local council and government facility, or health facility characteristics (such as type of water source, availability of drinking water at the facility, whether a separate maternity unit is available, electricity shortages). In one out of five measures of monthly supply of drugs (i.e., Quinine), the treatment group, on average, has a marginally higher supply in the year prior to treatment. In one out of four user-charge measures, there is some evidence (the estimate is significant at the 10 percent level) that patients served by the treatment facilities are more likely to pay for service delivery. The control group also appears to have fewer unqualified workers (staff with less than advanced A-level education), although the difference is only significant at the 10 percent level. [15]

We also calculated average standardized pre-treatment effects by estimating (5) on each family of outcomes (utilization, utilization pattern, quality measures, catchment area statistics, health facility characteristics, citizen perceptions, supply of drugs and funding, and user charges) using pre-intervention data. We cannot reject the null hypotheses of no difference between treatment and control groups (table 1c). Thus, overall the randomization appears to have been successful.

5.2 Processes

The initial phase of the project, i.e., the three separate meetings, followed a pre-design structure. A parallel system (a member of the survey team originating from the district participated as part of CBO team during the meetings) also confirmed that this initial phase of the intervention was properly implemented. After these initial meetings, it was up to the community to sustain and lead the process. In this section, we present some evidence on this first component in the accountability chain depicted in figure 2; namely if the treatment communities become more involved in monitoring the providers.

To avoid influencing local initiatives, we did not have any external agents visiting the communities during the study period. Therefore, we are not able to document

[15] As discussed in sub-section 4.3, we control for these chance differences between groups in the distribution of pre-assignment characteristics.

all actions taken by the communities in response to the intervention. Still, we have some information on how processes in the community have changed. Specifically, the CBOs submitted reports on what type of changes they observed. We use facility and household survey data to corroborate these reports.

According to the CBO reports, the community-based monitoring process that followed the first set of meetings was a joint effort mainly managed by the local councils, HUMC (Health Unit Management Committee) and community members. In the treatment communities, the performance of the health facility was discussed during village meetings. The local council survey also confirms this. A typical village in the treatment group had, on average, six local council meetings in 2005. In those meetings, 89 percent of the villages discussed issues concerning the project health facility. The main subject of discussion in the villages concerned the community contract or parts of it, such as behavior of the staff.

The CBOs report that concerns raised by the village members were carried forward by the local council to the health facility or the HUMC. However, although the HUMC is an entity that should play an important role in monitoring the provider, it was in many cases viewed as being ineffective. As a result, mismanaged HUMCs were dissolved and new members elected, while others felt the pressure from the community to act and follow up on the issues covered in the community contract. These claims are also confirmed in the survey data: More than one third of the HUMCs in the treatment communities were dissolved and new elected or received new members following the intervention. In the control communities, we observe no dissolved HUMC. Further, the CBOs report that the community, or individual members, also monitored the health facility staff during health visits to the clinic, when they rewarded and questioned issues in the community contract, which had or had not been addressed, suggesting a more systematic use of non-pecuniary rewards. Tools such as suggestion boxes (where community members could anonymously leave suggestions for change or comment on the lack of change that was supposed to have taken place), numbered waiting cards (to ensure a first-come-first serve basis), and duty roasters, were also reported to be put in place in several treatment facilities.

In table 2 we formally look at the program impact on these processes.[16] We use data collected through visual checks by enumerators during the post-intervention survey. As reported in table 2 (specifications 1-2), one year into the project, treatment facilities are significantly more likely to have suggestion boxes (no control facility had these, while 36 % of the treatment facilities did) and numbered waiting cards (only one control facility had these, while 25 % of the treatment facilities did). A higher share of treatment facilities also post information on free-services and patient's rights and obligations (specifications 3-4). The enumerators could visually confirm that 70 percent (17 out of 25) of the treatment facilities had at least one of these "monitoring tools" (suggestion boxes, numbered waiting cards, and/or posters on free-services), while only 4 out of 25 control units had at least one of them. The difference is statistically highly significant (specification 6). Column (5) reports the average standardized effect of the

[16] We did not collect data on these processes in the pre-intervention survey.

monitoring tools. The estimate is significantly different from zero at the 1 percent level.

The results based on household data mirror the findings reported in columns 1-6, table 2. For example, the performance of the staff is more often discussed in local council meetings in the treatment communities (specifications 7), suggesting that the treatment communities became more engaged. Three out of four households surveyed in the treatment villages have attended at least one village meeting in 2005. Of those attending, 40 percent (13 percentage points) more households in the treatment community report that the functioning of the health facility was discussed. Combining the evidence from the CBO reports and the household survey data thus suggests that both the "quantity" of discussions about the project facility and the subject (from general to specific discussions about the community contract) changed in response to the intervention.[17]

5.3 Treatment practices

The qualitative evidence from the CBOs and, to the extent that we can measure it, the findings reported in table 2, confirm that the treatment communities became more involved in monitoring the provider. Did community-monitoring affect the health worker's behavior and performance? We turn to this next. We report the results on treatment practices and staff behavior using quantitative indicators such as the immunization of children, waiting time, staff absenteeism, examination procedures, management of the clinic, drug leakage, and extent of preventive care.[18]

We start by looking at examination procedures.[19] Regression 1, table 3, presents the result of estimating (4) with the dependent variable being an indicator of whether any equipment (for instance thermometer or blood pressure equipment) was used during the examination. 49 percent of the patients in the treatment community reported that equipment was used the last time the respondent (or the respondent's child) visited the project clinic, as opposed to only 41 percent in the control group. The difference-in-differences estimate, 8 percentage points or a 19% increase, is highly significant.

In regression 2, table 3, we look at an alternative measure of staff performance – the waiting time – defined as the difference between the time the user left the facility and the time the user arrived at the facility, subtracting the examination time. On

[17] Additional evidence on community engagement and monitoring is reported in the working paper version.

[18] We report the results on treatment practices and staff behavior using perception responses by households in the working paper. The perception results corroborates the findings reported above.

[19] Naturally, the relevant treatment is conditional on illness and the condition of the patient. However, since the project was randomly allocated across communities, there is no reason to believe that the type of illness and the condition of the patients should differ systematically across groups. In fact, we have information on reported symptoms for which the patient seeks care (from the household survey). There are, on average, no systematic differences in reported symptoms across treatment and control communities.

average, the waiting time was 133 minutes in the control facilities and 119 in the treatment facilities. The difference is significant.[20]

Table 4, column 1, reports the results on absenteeism.[21] The point estimates suggest a substantial treatment effect. On average, the absence rate, defined as the ratio of workers not physically present at the time of the post-intervention survey to the number of workers employed, is 14 percentage points lower in the treatment facilities. Column 2 presents the result when only using the nominator as the dependent variable. In the treatment facilities, the median number of workers present was 3 as compared to 2 in the control clinics. Thus, in response to more extensive community monitoring, health workers are more likely to be at work.[22]

The findings on immunization of children under five are reported in tables 5a-5f.[23] We have information on how many times (doses) in total each child has been immunized with polio, DPT, BCG, and measles. To the extent that this is possible, these data were collected from households' immunization cards.

According to the Uganda National Expanded Program on Immunization (UNEPI), each child in Uganda is suppose to be immunized against measles (one dose at 9 months and two doses in case of an epidemic); DPT (three doses at 6 weeks, 10 weeks and 14 weeks); BCG (one dose at birth or during the first contact with a health facility); and polio (three doses, or four if delivery takes place at the facility, at 6 weeks, 10 weeks and 14 weeks). To account for these immunization requirements, we create dummy variables taking the value of one if child i of cohort (age) j had

[20] The point estimates for the treatment effect in table 3 are similar, but somewhat less precisely estimated, when only using data from the post-intervention survey, i.e. when estimating (2) instead of (4).

[21] The post-intervention survey was not announced in advance. At the start of the survey, enumerators physically verified the provider's presence. A worker was counted as absent if, at the time of the visit (during facility hours), he or she was not in the clinic. Staff reported to be on outreach were omitted from the absence calculation. Absence rate is the ratio of workers not physically present at the time of the post-intervention survey to the number of workers on the list of employees in as reported in the pre-intervention survey. Four observations were dropped because the total number of workers verified to be present or reported to be on outreach exceeded the total number of workers on the pre-intervention staff list. Assuming instead no absenteeism in these four facilities yields a point estimate of -0.20 (0.065). In the full sample, 47 percent of the health workers were absent. Chaudhury et al. (2006), based on a larger sample of both rural and urban health centers in Uganda, report that 37 percent of the workers, on average, are absent.

[22] As reported in the working paper, enumerators also visually checked the condition of the health center, i.e. whether floors and walls were clean, the condition of the furniture and the smell of the facility. We combine these variables through principal components analysis into a summary score. The point estimate implies that treatment clinics, on average, score 0.56 standard deviations higher than the control facilities. Thus, treatment clinics appear to have put more effort into keeping the clinic in decent condition in response to the intervention. Improvements in treatment practices are also substantiated by the qualitative data assembled.

[23] We report results of estimating (2) rather than the difference-in-differences equation (4), since the pre-treatment vaccination outcomes were strongly influenced by a mass immunization campaign implemented prior to the survey period. Due to reported irregularities in the top management of the unit in charge of the immunization campaigns, we have not been able to assemble accurate information on the actual timing of the campaign prior to the intervention.

received the required dose(s) of measles, DPT, BCG, and polio, respectively, and zero otherwise. We then estimate (5), using the binary indicators (for measles, DPT, BCG, and polio) as dependent variables for each age group (newborns (under 3 months), under 1 year, 13-24 months, 25-36 months, 37-48 months, and 49-60 months). The results are reported in tables 5a-5f.[24]

The summary measures, i.e., the average standardized effects, are significantly positive for the younger cohorts (tables 5a-5c). Looking at individual effects, there are significant positive differences between households in the treatment and control community for all four vaccines, although not for all cohorts. For example, the program impact on measles vaccinations for one-year old children is positive and significant (table 5c). In the control group, 79 percent of the children have been immunized, while the corresponding number in the treatment group is 5 percentage points higher. Overall, the effects are larger the younger the cohort. For example, twice as many newborns in the treatment group have received Vitamin A supplement, 43% more newborns have received the first dose of BCG vaccine, and 40% more newborns have received the first dose of polio vaccine as compared to the control group.

According to the government health sector strategic plan, preventive care is one of the core tasks for health providers at the primary level. Although we did not collect data on households' knowledge about health and various preventive measures, we have data on to what extent households have been informed about the potential dangers of self-treatment and if they have received information about family planning. Table 4 shows that a significantly larger share of households in the treatment communities has received information about the dangers of self-treatment (specification 3), and the importance of family planning (specification 4). The difference is 7 and 6 percentage points, respectively.[25]

In the working paper version we also document indirect evidence of reduced drug leakage. As reported in section 5.7, we do not find any systematic difference in supply of drugs (from public agencies) between the treatment and control groups. However, stock-outs of key drugs are reported to be occurring at a higher frequency in the control facilities even though, as reported in the next section, the control facilities treat significantly fewer patients. These findings suggest that more drugs leaked out of facilities lacking community monitoring.

[24] The first dose of DPT, BCG, polio, and Vitamin A supplement should be given at birth or in the first couple of months after delivery. Measles vaccination is an exception and should not be given at birth. Since first dose should be given at 9 months, we exclude immunization against measles for infants under 12 months.

[25] As a reference point, the share of households that have received information about the dangers of self-treatment and the importance of family planning are 32 percent and 30 percent, respectively, in the control communities, implying a 28% and 23% increase in health knowledge.

5.4 Utilization

The evidence presented so far shows that treatment communities began to monitor the health unit more extensively in response to the intervention and that in light of better community monitoring, the health unit staff responded by improving the provision of health services. We now turn to the question of whether increased community monitoring also resulted in improved quantity and quality of care.

We collected detailed data from the health facilities on the number of out-patients, the number of deliveries, the number of antenatal care patients, and the number of people seeking family planning services.[26] Table 6 presents the results, for the four different utilization variables, from the estimation of equation (5). The average (of the four utilization measures) standardized treatment effect is positive and highly significant. One year into the program, utilization (for general outpatient services) is 16 percent higher in the treatment facilities. When controlling for district fixed effects, the point estimate is slightly larger and more precisely estimated. The difference in the number of deliveries at the facility (albeit starting from a low level) is even larger (68 percent, regression 4) and fairly precisely estimated. There are also positive differences in the number of patients seeking antenatal care (22 percent, regression 8) and family planning (60 percent, regression 10), although these estimates are not individually significantly different from zero.

As a complement to the difference approach, columns 3 and 6 present the results from the estimation of a value added specification.[27,28] The point estimate on out-patients implies a 22 percent increase in average number of patients treated. The working paper version also reports difference-in-differences estimates, i.e. estimates of (4). The point estimates on out-patients and deliveries imply a 28 percent and 38 percent, respectively, increase in utilization, although these point estimates are not statistically different from those reported in specifications (1) and (4) in table 6 (but significantly different from zero).

Table 7 reports changes in utilization patterns based on household data. We collected data on where each household member sought care during the last year in case of illness that required treatment. Apart from recording visits to the project facility (treatment or control facility), we recorded visits to private providers (for-profit and NGOs), traditional healers, self-treatment (i.e., purchases of medicine in drug shops), or other government facilities (i.e., not a project facility). Consistent with the findings reported in table 6, we find a positive and significant difference in the use of the

[26] As discussed in section 5, these data were assembled by counting the number of patients from daily patient records, maternity unit records, the antenatal care register, and the family planning register.

[27] Data on the number of antenatal care patients and the number of people seeking family planning services were not collected from medical records in the pre-treatment survey.

[28] The value added specification is

$$y_{jdPOST} = \alpha_{VA} + \beta_{VA} T_{jd} + \lambda y_{jdPRE} + \varepsilon_{jdPOST} \ .$$

project facility between treatment and control facilities (specification 1). The increase, 14 percent higher in the treatment group as compared to the control group, is similar to that reported in table 6 (using facility records).

Table 7 also shows that households in the treatment community reduced the number of visits to traditional healers and the extent of self-treatment (specification 6), while there are no statistically significant differences (regressions 2, 3, and 7) across the two groups in the use of other providers (NGO, for-profit, or other government facilities). Thus, households in the treatment communities switched from traditional healers and self-treatment to the project facility in response to the intervention.

5.5 Health outcomes

The main objective of the community-based monitoring project was to improve health outcomes in rural areas of Uganda where health indicators have been stagnating. To achieve this objective the project intended to enhance communities' abilities to monitor the public health care provider, thereby strengthening providers' incentives to increase both the quality and quantity of primary health care provision. As reported above, the project was successful in raising both utilization and, to the extent that this can be measured, service quality. Next, we turn to health outcomes.

Data on two health outcomes were collected. First, we collected information on whether the household had suffered from the death of a child (under five years) in 2005, i.e., the first year of the community-monitoring project. Second, we measured the weight of all infants (i.e., under 18 months of age) and children (between 18 and 36 months of age) in the surveyed households.[29]

Health outcomes (under-five mortality and weight of infants) could have improved for several reasons. As noted in the Introduction, access to a small set of proven, inexpensive services could, worldwide, have prevented more than half of all under-five deaths. For a country with an epidemiological profile as in Uganda, the estimate of preventable deaths is 73% (Jones, et al. 2003).[30] In the community monitoring project specifically, increased utilization and having patients switching from self-treatment or traditional healers to seeking care at the treatment facility could have an effect. Holding utilization constant, better service quality, increased immunization, and more extensive use of preventive care (health education) could also result in a reduction in mortality and improved health status.

[29] The weighing scale was a regular hanging baby scale with trousers (Salter type). Two trained enumerators assisted in the task. During the weighing process, the enumerators took help from family members, mostly mothers. When the infant/child was hanging calmly on the scale, the enumerators recorded the weight.

[30] This is likely to be a conservative number since only medical interventions for which cause-specific evidence of effect was available were included in the estimation. For example, increased birth spacing (as a result of family planning), which has been estimated to reduce under-5 mortality by 19% in India was not considered. Moreover, several perinatal and neonatal health interventions that could be implemented in low-income countries were not included (Darmstadt et al, 2005).

As a reference point we review the set of health intervention feasible for delivery at high coverage in low-income settings with sufficient evidence of effect on reducing mortality from the major causes of under-five deaths (Jones et al, 2003). We focus on community-based, randomized, controlled field trials that bear some resemblance (because they are community-based) to our project. Several of these field trials document reductions in under-five mortality rates of 30-50% one to two years into the project.[31] There is, however, a fundamental difference between the interventions discussed in footnote (31) and our work. The medical field trials address the question of impact of a biological agent(s) or treatment practice(s) in a community setting *when the community health workers and/or medical personnel competently carry out their tasks.*[32] Here on the contrary, we do not introduce new health interventions or increase supply of health inputs, but instead we focus on *incentivizing health workers to competently carry out their tasks through strengthened local accountability.*

Table 8 presents the results on child mortality. 3.2 percent of the surveyed households in the treatment community had suffered from the death of a child in 2005. The corresponding number in the control community is 4.9 percent, implying a 1.7 percentage points, or 33 percent, reduction in child deaths in the treatment communities. Thus, our non-medical intervention compares favorably to some of the more successful community based intervention trials reported in the medical literature (see footnote 31). While the effect is large, it is worth emphasizing that the 90 percent confidence interval of our estimate also includes much lower effects (90% CI: 0.3%-3.0%). With a

[31] For example, a project in Tigray, Ethiopia, in which mother coordinators, supported by a team of supervisors, were trained to teach other local mothers to recognize symptoms of malaria in their children and provide antimalarials, reduced under-5 mortality by 40% (Kidane and Morrow, 2000). Bang et al. (1990) document a 30% reduction in under-five mortality from an intervention including mass education about childhood pneumonia, and case management of pneumonia (treatment with antibiotics) by trained village health workers and traditional birth attendants — a result similar to the meta-analysis estimate of Sazawal and Black (2003). Bang et al. (1999) evaluate a project in which trained village health workers, assisted by birth attendants and supervisory visits, provided home-based neonatal care, including treatment of sepsis. Two years into the project they document a reduction in infant mortality by nearly 50%. Rahmathullah, et al. (2003) assess the impact of a community-based project in two rural districts of Tamil Nadu, India, where newborn infants in the treatment group were allocated oral vitamin A after delivery. The intervention resulted in a 22% reduction in total mortality at age 6 months. Manandhar et al. (2004) evaluate a project in which a facilitator convened nine women's group meeting every month in the Makwanpur district in Nepal. The facilitator further supported groups in identifying perinatal problems and formulated strategies to address them. Two year into the project, they document a 30% reduction in neonatal mortality. As part of a field trial of cholera vaccine in rural Bangladesh, Rahman et al. (1982) evaluate the impact of immunization of women with tetanus injections during pregnancy. The intervention reduced neonatal mortality by 45%. Mtango and Neuvians (1986) evaluate a project in rural Tanzania in which trained village health workers visited families at their homes every six to eight weeks, giving health education on recognition and prevention of acute respiratory infections, treating children with pneumonia with antibiotics or referring them to the next higher level of care. Within a two-year period, they document a 27% reduction in under-five mortality – a reduction slightly lower than that found in a similar study in rural Bangladesh (Fauveau et al., 1992).

[32] In the medical field trials this is ensured through close supervision and support by the project evaluators throughout the study period.

total of approximately 55,000 households residing in the treatment communities, the treatment effect (0.017) corresponds to 546 averted under-five deaths in the treatment group in 2005.[33]

The dependent variable in regression 3, table 8, is estimated under-five mortality rate in the community.[34] Consistent with the findings in columns 1-2, the point estimate suggest a substantial treatment effect. The average under-five mortality rate in the control group is 145, close to the official figure of 133 for 2005 (UNICEF, 2006). In the treatment group, the under-five mortality rate is 97. The difference is significant (specification 3) and fairly precisely estimated when controlling for district fixed effects (specification 4).

The program impact on the weight of infants is reported in table 9. Growth charts for boys and girls are depicted in figure 4. As in Cortinovis et al's (1997) study of over 4,000 children from 31 villages in Mbarara (a district in south-western Uganda), we find that Ugandan infants have values of weight far lower than the NCHS/CDC international reference. The gap increases for older infants. The median weight of six-month old boys in the sample is close to the 25th percentile of the NCHS/CDC reference chart. For the 18 months old, the median weight for boys lies close to the 10the percentile of the NCHS/CDC chart.

Figure 5 plots the distribution of weight-for-age (z score).[35] A population similar to the reference population (NCHS) will have a mean z score of zero, with approximately 2.5 percent of the population below a z score of -2 (the threshold for moderately underweight). In the sample of measured infants, 17.4 percent fall below this threshold. 8.5 percent of the infants (up to 18 months) are severely underweight (< -3 z scores). Almost a quarter of the infants falls below the mildly underweight threshold (< -1 z score).

The difference in means of z scores of infants between the treatment and the control group is reported in regression 1, table 10. The estimated effect (difference) is 0.164 z score in weight-for-age. Regression 2 applies a more stringent restriction on the data to avoid problems of misreporting.[36] The difference in mean is 0.17 z score and is precisely estimated. Figure 6 plots the distribution of z scores for treatment and control

[33] We get an almost identical estimate (540 averted deaths) when we weight with distance to the health facility. Since villages closer to the facility were oversampled, the sample of treatment villages is not fully representative of the total population in the treatment communities.

[34] The under-five mortality rate is estimated as the number deaths of children under five in the community as a fraction of number of live births in 2005 (i.e. number of infants younger than one year at the end of 2005 plus the number of infants under one year that died in 2005) expressed per 1,000 live births.

[35] The z-score is a normally distributed measure of growth defined as the difference between the weight of an individual and the median value of weight for the reference population (2000 CDC Growth Reference in the U.S.) for the same age, divided by the standard deviation of the reference population. We exclude z scores $> |4.5|$ as implausible. Four observations (out of 1142) with z scores < -4.5 were consequently dropped.

[36] Specifically, we drop observations with a recorded weight above the 90th percentile in the growth chart reported in Cortinovis et al (1997). Since trained enumerators measured weight, the reporting error is likely due to misreported age of the child.

groups. The difference in measured weight is most apparent for underweight children. Underweight status causes a decrease in immune and non-immune host defenses. Thus, since underweight children are at a higher risk of suffering from infectious diseases (and more severe complications of infectious diseases), and therefore in higher demand for/need of health care, the data in figure 6 are consistent with a positive treatment effect arising from improved access and quality of health care, rather than a general increase in nutritional status.

Regression 3 adds district fixed effects and controls for age and gender. The results remain qualitatively unchanged. The incidence of underweight increases with age. We cannot reject the hypothesis that the treatment effect is the same for girls and boys.

The treatment effect is quantitatively important. For this purpose, the baseline proportion of infants in each risk category (severe, < -3 z scores; moderately, $-3 \leq z$ scores < -2; mild, $-2 \leq z$ scores < -1) in the control group was calculated. Applying the shift in the weight-for-age distribution (adding 0.17 z score) with the odds ratio for each category – children who are mildly [moderately] {severely} underweight have about a two-fold [five-fold] {eight-fold} higher risk of deaths from infectious disease (Jones et al., 2003) – the reduction in average risk of mortality is estimated to be approximately 8 percent (figure 6).[37]

Columns 4-5 in table 9 report the program impact on child weight for children between 18-36 months of age. The treatment effect is small and insignificant.[38]

5.6 Robustness

Given that within each district there are both treatment and control units, one concern with the evaluation design is the possibility of spillovers from one catchment area to another. For example, if a treatment facility improved the quality of health provision due to the intervention, households in villages in the catchment area of a control community might choose to seek service in the treatment facility. If this is the case, we would overestimate the effects (on utilization) of the intervention. It is also possible that community members in the control group copied the monitoring approach of the treatment group, in which case the bias would go in the opposite direction.

In practice, there are reasons to believe that this is not a serious concern. First, the average (and median) distance between the treatment and control facility is 30 kilometers. Second, in a rural setting, it is unclear to what extent information about improvements in treatment facilities has spread to control communities. Still, the

[37] To put this into perspective, a review of controlled trials designed to improve the intake of complementary food for children aged six months to five years showed a mean increase of 0.35 z score (Jones et al, 2003). If the present coverage level were increased to universal coverage (99%), Jones et al estimate that complementary feeding alone would prevent 6% of the under-five deaths in the 42 countries with the 90% of worldwide child deaths in 2000. According to Jones et al, this is one of the most effective (in the sense of preventing under-five deaths) preventive interventions feasible for delivery at high coverage in a low-income setting.

[38] Measurement errors due to misreported age of the child are likely to be a more serious concern for children above 18 months than for infants.

possibility of spillovers is a concern. One way of testing for spillover effects is to estimate an augmented version of (2) for the sample of control facilities.[39] That is, we estimate

$$y_{idPOST} = \alpha + \eta DIST_{id} + \varepsilon_{idPOST}, \tag{6}$$

where $DIST_i$ is the distance (in kilometers) between the control facility i and the closest treatment facility. The results of estimating (6) for the various utilizations measures are reported in table 10.[40] In all specifications, the estimate of η differs insignificantly from zero.

Another concern, which does not influence the casual effect of the project but the interpretation, is if the district or sub-district management changed their behavior or support in response to the intervention. For example, the Health Sub-district or local government may have provided additional funding or other support to the treatment facilities. The results in table 11 do not provide any evidence of this being the case. Difference-in-differences estimates of the monthly supply of drugs indicate that the treatment and control facilities are similar. If anything, drug supplies are smaller in the treatment clinics. The treatment facilities did not receive more funding from the sub-district or district (regression 6) as compared to the control facilities. The difference-in-differences estimate is negative, but insignificant. As reported in the working paper, we do not find any differences in constructions or infrastructure during the first project year or in availability of equipment at the health facility.

Upper-level authorities could also have increased their supervision and control of treatment facilities in response to the intervention. However, this does not seem to be the case either. The supervision of providers by upper-level government authorities remained low in both the treatment and the control group (table 12, regressions 1-2).

The incidence of supervision and control visits may be an imprecise measure of the effectiveness of monitoring by the upper-level authorities. A complementary measure is implemented sanctions. We have data on the extent to which staff was dismissed or transferred during the first year of the project. As noted in section 4, only the District Service Commission has the authority to dismiss and transfer staff. There is only a handful of staff that has been dismissed or transferred in 2005 and there is no systematic pattern that distinguishes treatment from control facilities (table 12, regressions 3-4). Likewise, there is no difference between treatment and control facilities in the number of staff that voluntarily left the facility during 2005 (specification 5) or in the number of new workers in 2005 (specification 6).

Yet another interpretational concern is that the intervention directly influenced health workers' behavior. For example, it is possible that the workers decided to exert higher effort into serving the community as a result of the health facility staff meeting (where information on outcomes was disseminated). Under this alternative

[39] Pooling the sample of control and treatment facilities and adding a dummy for treatment facilities yields identical results.

[40] A difference-in-differences version of (6) is reported in the working paper. The results mirror those in table 12.

hypothesis, the more engaged community, as is evident from the results reported in table 2, is an inconsequential by-product of either the intervention or changed staff behavior. However, since we do not observe any changes in the degree of supervision and control by upper-level government authorities (remained low in both the treatment and the control group), and since we did not have any external agents visiting the communities or health facilities during the period of study, it is difficult to see how a couple of meetings would have changed the health workers' incentives to such extent that we one year after observe large differences in both quality and quantity of service provision. This alternative explanation also is also at odds with the CBO reports (and the corroborating evidence presented in table 2), which stress that the intervention initiated a process led by various actors in the community (local councils, Health Unit Management Committees, and individual community members). These actors, when having access to relevant information and when being able to coordinate on expected reforms, should be in a position to monitor effort and thereby choose to more systematically exploit the instruments (social rewards and sanctions) at their disposal. This way they should have been able to affect the health workers' incentives.[41]

Taken together, these findings reinforce our confidence that the improved quality and quantity of health care provision resulted from increased efforts by the health unit staff to serve the community in light of better community monitoring.

6 Discussion

Based on a small but rigorous empirical literature on community participation and oversight, and extensive piloting in the field, our conjecture was that lack of relevant information and failure to agree on, or coordinate expectations of, what is reasonable to demand from the provider were holding back individual and group action to pressure and monitor the provider. We designed an intervention aimed at relaxing these constraints. Through two rounds of community meetings, local NGOs initiated a process aimed at energizing the community and agreeing on an action plan.

The intervention managed to rejuvenate (formal and informal) institutions of beneficiary control. We document large increases in utilization and improved health outcomes that compare favorably to some of the more successful community-based intervention trials reported in the medical literature. The intervention we evaluate, however, differ in at least one important dimension from the medical field trials. Specifically, while they address the question of impact of a biological agent(s) or treatment practice(s) when the community health workers and/or medical personnel do what they are suppose to do, we focus on a mechanism to ensure that health workers exert effort to serve the community.

[41] We pondered the idea of adding an additional treatment of only health staff meetings but decided against the idea for both financial and ethical reason (why withholding information to the community when the community is the actor that could use it to put pressure on the provider).

The Citizen report card project was implemented in nine different districts of Uganda and reached approximately 55,000 households. Thus, in this dimension, the project has already shown that it can be brought to scale. However, the literature on how to enhance local accountability is still in its infancy so there are still a number of outstanding questions to be answered before concluding that more funding should be directed to various activities with the aim of strengthening beneficiary control. For example, we know little about long-term effects. We know little about cross-sector externalities, i.e. it is possible that the treatment communities' ability to coordinate citizen actions has also been applied to other areas of concern, in which case the aggregate returns are higher than what they appear from the results reported here. It may also be the case that combining bottom-up monitoring with a reformed top-down approach yields even better results.

Before scaling up, it is also important to subject the project to a cost-benefit analysis and relate the cost-benefit outcomes to other possible interventions. This would require putting a value on the improvements we have documented. To provide a flavor of such a cost-benefit analysis, consider the findings on averting the death of a child under five. The intervention resulted in 1.7 percentage points fewer child deaths in the treatment communities during the first project year. To the extent that this number is representative of the total treatment population, this would imply that approximately 550 under-five deaths were averted as a result of the intervention. A back-of-the-envelope calculation then suggests that the intervention, only judged on the cost per death averted, must be considered to be fairly cost-effective. The estimated cost of averting the death of a child under five is around $300 in the Citizen report card project. This can be compared to the numbers reported by Filmer and Pritchett (1999). They contrast the cost of averting the death of a child derived from increasing public expenditures on health (regression estimates range from $47,112 to $100,927), to more conventional health interventions based on cost-effectiveness estimates of the minimum required cost to avert a death (ranges from $1,000 to $10,000 for diarrheal diseases, from $379 to $1,610 for acute respiratory infection, $78 to $990 for malaria, and $836-$3,967 for complications of pregnancy). Bryce at al. (2005) estimate that the average cost per child life saved through the combined and integrated delivery of 23 interventions shown to reduce mortality from the major causes of death in children younger than 5 years is $887.[42]

Although research on medical interventions is plentiful, little is known about the characteristics of delivery strategies capable of achieving and maintaining high quality and coverage. As argued in a recent *Lancet* article, a systematic program of research

[42] These numbers should be viewed with caution. Naturally, the 95 percent confidence interval would also include a much smaller estimate of program impact than the 1.7 percentage points used here. Moreover, since the largest cost item was the collection of data and these data were partly used in the intervention and partly to evaluate impact, the cost is a rough estimate. Filmer and Pritchett's (1999) estimates of the cost of averting a child death derived from increasing public expenditures on health are subject to a variety of estimation problems and the health interventions based cost-effectiveness estimates of the minimum required cost to avert a death are, as noted by Filmer and Pritchett, at best suggestive.

to answer questions about how best to deliver health (child survival) interventions therefore is urgently needed (Bryce, et al., 2003). Strengthening the relationship of accountability between health service providers and citizens has been put forward as one important component to improve access to and quality of health care. The Citizen report card project has shown that it can work. Future research should address long term effects, experiment with alternative tools, and study to which extent the results could be generalized to other social sectors.

References

Appleton, S., 2001, "The Rich Are Just Like Us, Only Richer': Poverty Functions or Consumption Functions?", *Journal of African Economies* 10(4): 433-469.

Banerjee, A. and E. Duflo, 2005, "Addressing Absence", *Journal of Economic Perspectives* 20 (1): 117-132.

Banerjee, A., A. Deaton and E. Duflo, 2004, "Wealth, Health, and Health Service Delivery in Rural Rajasthan", *American Economic Review Papers and Proceedings* 94(2): 326-330.

Banerjee, A. and R. He, 2003, "The World Bank of the Future", *American Economic Review* 93(2): 39-44.

Bang A.T., R.A. Bang, S.B. Baitule, M.H. Reddy, M. D. Deshmukh, 1999, "Effect of home-based neonatal care and management of sepsis on neonatal mortality: field trial in rural India", *The Lancet*, 354(9194): 1955-61.

Bang A.T., R.A. Bang, O. Tale, P. Sontakke, J. Solanki, R. Wargantiwar, and P. Kelzarkar, 1990, "Reduction in pneumonia mortality and total childhood mortality by means of community-based intervention trial in Gadchiroli, India", *The Lancet*, 336: 201-6.

Besley, T. and R. Burgess, 2002, "The Political Economy of Government Responsiveness: Theory and Evidence From India", *The Quarterly Journal of Economics* 177(4):1415-51.

Bjorkman, M., and J. Svensson, 2007, "Power to the People: Evidence from a Randomized Field Experiment of a Community-Based Monitoring Project in Uganda", CEPR Working Paper # 6344.

Chaudhury, N., J. Hammer, M. Kremer, K. Muralidharan, and F.H. Rogers, 2006, "Missing in Action: Teacher and Health Worker Absence in Developing Countries", *Journal of Economic Perspectives* 20(1): 91-116.

Duflo, E., R. Glennerster, and M. Kremer, 2006, "Using Randomization in Development Economics Research: A Toolkit", CEPR Discussion Paper No. 6059.

Duflo, E. and R. Hanna, 2005, "Monitoring Works: Getting Teachers to Come to School", Working Paper, Department of Economics and Poverty Action Lab, MIT.

Fauveau V., M. K. Stewart, J. Chakraborty, S. A. Khan, 1992, "Impact on mortality of a community-based programme to control acute lower respiratory tract infections" *Bulletin of the World Health Organization*, 70(1): 109-16.

Filmer, D. and L. Pritchett, 1999, "The Impact of Public Spending on Health: Does Money Matter?", *Social Science and Medicine* 49(10).

Khemani, S. 2006, "Can Information Campaigns Overcome Political Obstacles to Serving the Poor," mimeo, The World Bank.

Kling, J.R., J. Liebman, L.F. Katz and L. Sanbonmatsu, 2004. "Moving To Opportunity and Tranquility: Neighborhood Effects on Adult Economic Self-Sufficiency and Health from a Randomized Housing Voucher Experiment," Working Papers 5, Princeton University, Department of Economics, Industrial Relations Section.

Kidane, G. and R. Morrow, 2000, "Teaching mothers to provide home treatment of malaria in Tigray, Ethiopia: a randomised trial", *The Lancet*, 356(9229): 550-555.

Manandhar, D., D. Osrin, B. Shrestha, N. Mesko, J. Morrison, K. Tumbahangphe, S. Tamang, S. Thapa, D. Shrestha, B. Thapa, 2004, "Effect of a participatory intervention with women's groups on birth outcomes in Nepal: cluster-randomised controlled trial", *The Lancet*, 364 (9438): 970-979.

Mtango F.D. and D. Neuvians, 1986, "Acute respiratory infections in children under five years. Control project in Bagamoyo District, Tanzania", *Transactions of the Royal Society of Tropical Medicine and Hygiene*, 80(6):851-8.

McPake, B., D. Asiimwe, F. Mwesigye, M. Ofumbi, L. Ortenblad, P. Streefland and A. Turinde, 1999, "The Economic Behavior of Health Workers in Uganda: Implications for Quality and Accessibility of Public Health Services," *Social Science and Medicine* 49(7): 849-865.

Olken, B. 2005, " Monitoring Corruption: Evidence from a Field Experiment in Indonesia", NBER Working Paper No.11753.

Rahman, M., L.C. Chen, J. Chakraborty, M. Yunus, A.I. Chowdhury, A.M. Sarder, S. Bhatia, G.T. Curlin, 1982, "Use of tetanus toxoid for the prevention of neonatal tetanus. 1. Reduction of neonatal mortality by immunization of non-pregnant and pregnant women in rural Bangladesh", *Bulletin of the World Health Organization*, 60(2): 261-7.

Rahmathullah L, J.M. Tielsch, R.D. Thulasiraj, J. Katz, C. Coles, S. Devi, R. John, K. Prakash, A.V. Sadanand, N. Edwin, and C. Kamaraj, 2003, *BMJ*, 327(7409): 254.

Reinikka, R. and J. Svensson, 2004, "Local Capture: Evidence from a Central Government Transfer Program in Uganda", *The Quarterly Journal of Economics* 119 (2): 679-705.

Reinikka, R. and J. Svensson, 2006, "The Returns from Reducing Corruption: Evidence from Education in Uganda", Working Paper, IIES.

Republic of Uganda, 2000, "National Health Policy and Health Sector Strategic Plan 2000/01-2004/05", Ministry of Health, Kampala

Republic of Uganda, 2002, "Infant Mortality in Uganda 1995-2000: Why the Non-Improvement?", Discussion Paper No. 6, Planning and Economic Development, Kampala.

Singh, J. and M. Shah, 2002, "Community Score Cards in rural Malawi", World Bank, Washington, D.C. Processed

Strömberg, D., 2004, "Radio's Impact on Public Spending", *The Quarterly Journal of Economics* 119(1): 189-221.

Samuel P., 2002, *Holding the State to Account: Citizen Monitoring in Action*, Books for Change, Bangalore.

Sazawal S., R.E. Black, Pneumonia Case Management Trials Group, 2003, "Effect of pneumonia case management on mortality in neonates, infants, and preschool children: a meta-analysis of community-based trials", *The Lancet Infectious Diseases*, 3(9): 547-556.

UNICEF, 2006, *The State of the World's Children 2007: Women and Children - The double dividend of gender equality*, UNICEF.

UNDP, 2002, *Millennium Development Goals: A compact among nations to end human poverty*, Human Development Report 2003, UNDP.

World Bank, 2003, *Making Service Work for the Poor People*, World Development Report 2004, World Bank and Oxford University Press.

World Bank, 1996, "World Bank Participation Sourcebook", Environmental Department papers No. 19, Washington.

World Health Organization, 2006, *World Health Statistics 2006*, WHO Press, Switzerland.

Yamey G., on Behalf of the Interviewees, 2007, "Which Single Intervention Would Do the Most to Improve the Health of Those Living on Less Than $1 Per Day?" *PLoS Medicine* 4(10): e303.

A Appendix

A.1 Participatory Methods

The report card was disseminated to the community using a Participatory Rural Appraisal (PRA) methodology. In the early 1990s, the participatory rural appraisal methodology was mainly used by non-government organizations in East-Africa and South-Asia but are today widely used in many different organizations all over the world.[43] Participatory rural appraisal evolved from a set of informal techniques used by development practitioners in rural areas to collect and analyze data. It emphasizes local knowledge and the importance of having beneficiaries making their own appraisal, analysis, and plans for monitoring and evaluation of service providers. It is a participatory process intended to mitigate the collective action problem by facilitating the analysis of people's environment and identification and discussion of problems. The method employs a wide range of tools and techniques such as maps, diagrams, role-plays and action planning. Next, we briefly describe the specific tools used in the Citizen Report Card project in Uganda.

Venn diagrams were used to discuss power issues in service delivery. Participants were asked to list the different stakeholders in health service delivery (i.e. health facility staff, citizens, health management committee, district officials etc). Thereafter, the participants discussed the different roles and responsibilities of these players in ensuring the quality of the service, i.e. who is accountable to whom; what is a particular stakeholder accountable for, and how can these actors account for their actions. The outcome was used in the interface meeting to identify the stakeholders who have the power to ensure that quality service is delivered. The outcome also contributed to the process of developing a shared vision of how to monitor the provider.

Focus group discussions were used to generate discussions among and across subgroups. Participants were divided into key social groups such as women, men, youths, disabled, local leaders and elderly in order to get their perspectives over issues concerning service delivery and determine their preferences for change. Each group individually discussed the issues covered in the report card and recorded suggestions for improvements. Thereafter, each group presented the results to the other participants by using flip charts. In this way, the voice and priorities of all social groups were taken into considerations.

"Now, Soon, Later" approach is a technique aimed at helping the community identify issues they would like to address in the short term and those they would address in the longer term, considering the resource envelope at hand. Thereafter, the participants were asked to prioritize the needs according to their resource envelope and discuss which factors are important and necessary for making a change. This tool was intended to help the community analyze the resources available, the time frame for implementing the desired change and the importance of the issue.

Role-play was used to illustrate community and health facility interactions as per-

[43] See World Bank (1996).

ceived by the respective parties. This tool facilitated the discussion and dialogue in the interface meeting between health staff and community members. The story of the play illustrated the participants' interpretation of an ordinary day at the health facility. In the play, community members were asked to act the roles of health facility staff (In-charge; Mid-wife; Records Assistant; Watch Man; Laboratory Assistant; Senior Nurse etc) and health facility staff acted the roles of users of the facility (pregnant women; patients; poor patients; community leader; Chairman). Role-plays are viewed as an effective tool for diffusing sensitive issues (such as absenteeism or weak attention of staff). It is also a tool that can be used to illustrate constraints and opportunities, enabling users and providers to forge a way forward. Not only did the role play focus on the current situation at the health facility but in a second role play, the plot exemplified what the participants would like the situation to be like in six months.

Roles and Responsibility Analysis is used to provide clarity as to who is responsible for what activity. In this analysis, the participants review all planned activities in the action plan and ensure that each activity becomes someone's responsibility. This tool defines roles and responsibilities and helps strengthening the relationship of accountability between health service providers and citizens with regard to the activities determined in the action plan. The facilitator guides the participants to discuss the activities recorded in the action plan and help them agree on the criteria for taking up a responsibility for a particular activity. Thereafter, the participants identify who among the community or health facility staff would suit the criteria and discuss this responsibility with the person or group identified. The groups or individuals assigned to be responsible for a certain activity are then recorded in the action plan.

Action planning was a tool used in the final stage to summarize and record the community's suggestions for improvements (and how to reach them without additional resources). The action plan states the health issues/services that had been identified by the community and the staff as the most important to address; how these issues could be addressed; when they are supposed to be achieved; by whom this will be done; and how the community could monitor the improvements (or the lack thereof). The action plan is a contract between the community and the health facility. It forms the basis for local monitoring and makes it easier for the community to keep track of the implementation of agreed recommendations.

Table A.1. Total number of households, villages and enumeration areas in sample frame (50 units)

	Total	Within 1 km radius	Within 3 km radius excl. those within the 1 km radius	Within 5 km radius excl. those within the 3 km radius
Households	109,296	11,572	41,665	56,059
Villages	1,194	113	458	623
Enumeration areas	804			

Source: UBOS maps and census data

Table A.2. Sample frame characteristics (50 units)

	Mean	Median	Min	Max
Households in the catchment area	2,483	2,728	490	3,938
Households within 1 km radius in the catchment area	344	240	60	1014
Households within 3 km radius excl. those within the 1 km radius in the catchment area	1096	991	127	2,357
Households within 5 km radius excl. those within the 1 and 3 km radius in the catchment area	1,303	1,231	173	2,428
Villages in the catchment area	29	26	7	58
Villages within 1 km radius	3	3	1	8
Villages within 3 km radius excl. those within the 1 km radius in the catchment area	13	11	2	30
Villages within 5 km radius excl. those within the 1 and 3 km radius in the catchment area	15	15	2	31
Enumeration areas in the catchment area	20	19	4	35
Villages in enumeration area	1.9	2	0	6

Source: UBOS maps and census data.

Table A.3. Village characteristics in sample frame (50 units)

	Mean	Median	Min	Max
Number of households in village	92	84	0	273
Distance to facility	3.9	5	1	5

Source: UBOS maps and census data

Table A.4. Total number of households, villages in actual sample

	Total	Within 1 km radius	Within 3 km radius excl. those within the 1 km radius	Within 5 km radius excl. those within the 3 km radius
2004				
Households	4,978	1,239	2,024	1,715
Villages	293	70	121	102
2006				
Households	4,996	1,241	2,025	1,730
Villages	293	70	121	102

Table A.5. Village characteristics of actual sample

	Mean	Median	Min	Max
Number of households in village	102	92	22	232
Distance to facility	3.2	3	1	5

Figure 1: Timing of project

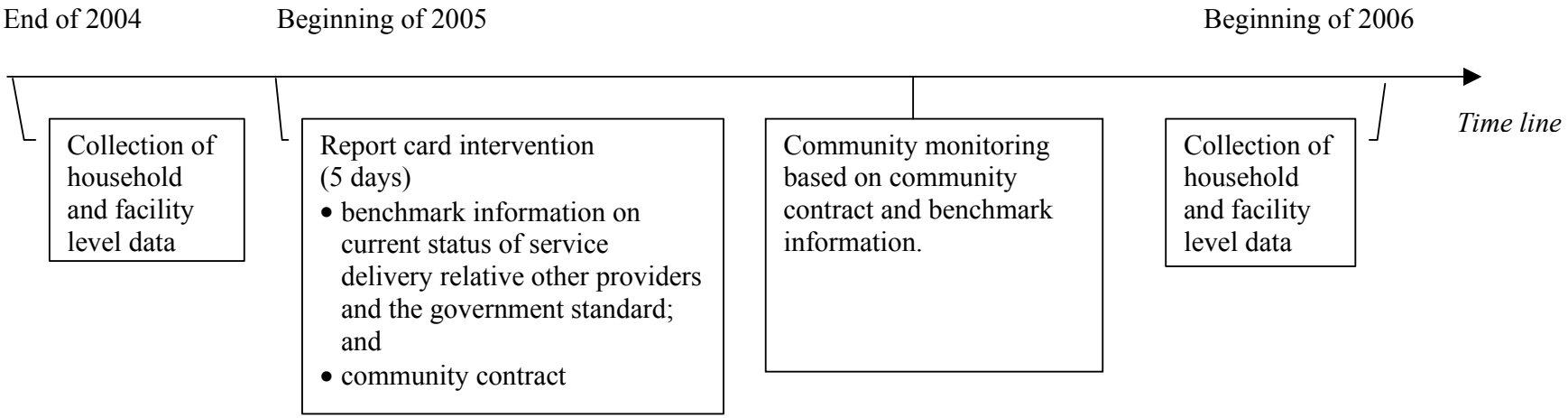


Figure 2: Schematic view of intervention and expected outcome

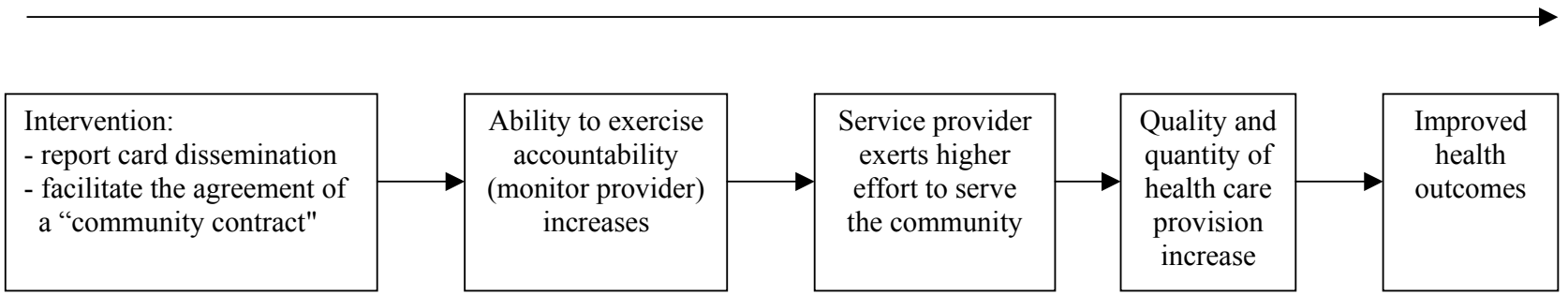


Figure 3: Alternative hypotheses

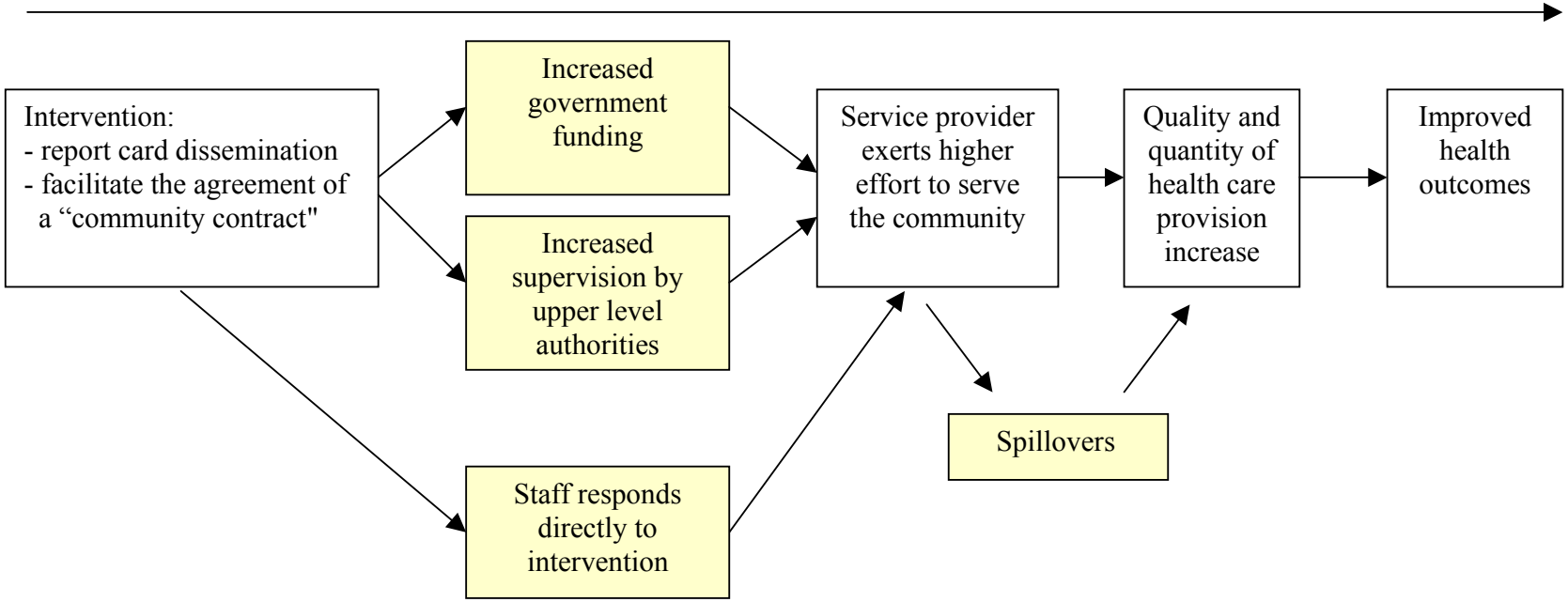


Table 1a. Average health facility and citizen characteristics, pre-treatment

Variables	Treatment group	Control group	Difference
Utilization:			
Out-patient care	593	675	-82
	(66)	(66)	(94)
Delivery	10.3	7.5	2.8
	(1.8)	(1.8)	(2.6)
Utilization pattern:			
Project facility	0.28	0.30	-0.03
	(0.02)	(0.02)	(0.03)
NGO facility	0.02	0.02	-0.002
	(0.01)	(0.00)	(0.007)
Private-for-Profit facility	0.22	0.24	-0.02
	(0.01)	(0.02)	(0.02)
Traditional healer	0.03	0.03	0.004
	(0.01)	(0.00)	(0.007)
Self treatment (drug shop)	0.35	0.31	0.046
	(0.03)	(0.02)	(0.032)
Other government facility	0.09	0.09	-0.007
	(0.01)	(0.01)	(0.015)
Other provider	0.01	0.01	0.007
	(0.00)	(0.00)	(0.005)
Quality measures:			
Waiting time	147	143	3.7
	(7.3)	(7.0)	(10.1)
Equipment usage	0.47	0.48	-0.01
	(0.04)	(0.04)	(0.06)
Funding at the facility:			
1000 shillings	4766	3429	1337
	(794)	(434)	(905)
Catchment area (CA) statistics:			
Villages in CA	23.2	24.6	-1.3
	(2.0)	(2.4)	(3.1)
Villages in CA – strata 1	2.64	1.84	0.80*
	(0.36)	(0.26)	(0.45)
Villages in CA – strata 3	8.88	9.52	-0.64
	(1.10)	(1.35)	(1.74)
Villages in CA – strata 5	11.7	13.2	-1.48
	(1.11)	(1.28)	(1.69)
Number of households in CA	2140	2224	-84.4
	(185)	(204)	(276)
Number of households per village	93.9	95.3	-1.42
	(5.27)	(6.32)	(8.23)

Notes: The results are catchment area (health facility) averages. Robust standard errors in parentheses. Significantly different from zero at 99 (***), 95 (**), and 90 (*) percent confidence level. Description of variables: See table 1c.

Table 1b. Average health facility and citizen characteristics, pre-treatment

Variables	Treatment group	Control group	Difference
Health facility characteristics:			
Piped water	0.04	0.04	-0.00
	(0.04)	(0.04)	(0.00)
Rain tank/Open well	0.52	0.36	0.16
	(0.10)	(0.10)	(0.14)
Borehole	0.44	0.60	-0.16
	(0.10)	(0.10)	(0.14)
Drinking water	1.76	1.48	0.28
	(0.17)	(0.12)	(0.20)
Drink safely today	0.40	0.32	0.08
	(0.50)	(0.48)	(0.14)
Separate maternity unit	0.16	0.16	0.00
	(0.07)	(0.07)	(0.10)
Distance to nearest Local Council I	0.72	0.85	-0.13
	(0.15)	(0.21)	(0.26)
Distance to nearest public health provider	9.78	11.1	-1.29
	(1.43)	(1.80)	(2.30)
Number of days without electricity in the last month	18.3	20.4	-2.12
	(2.95)	(2.90)	(4.14)
Number of staff with advanced A-level education	1.20	1.04	0.16
	(1.04)	(0.73)	(0.25)
Number of staff with less than advanced A-level education	4.84	3.68	1.16*
	(2.61)	(1.52)	(0.60)
Citizen perceptions:			
Polite behavior	1.04	1.03	0.01
	(0.01)	(0.01)	(0.01)
Attention	3.06	3.02	0.04
	(0.03)	(0.03)	(0.04)
Free to express	3.17	3.16	0.01)
	(0.03)	(0.03)	(0.04)
Citizens' informed about drug deliveries	0.13	0.16	-0.02
	(0.03)	(0.04)	(0.05)
Supply of drugs			
Erythromycin	420	346	74.2
	(111)	(71)	(131)
Chloroquine	3410	2915	495
	(456)	(338)	(567)
Septrine	2690	2430	260
	(476)	(403)	(623)
Quinine	573	335	238*
	(107)	(73)	(130)
Mebendazole	1597	1500	97
	(174)	(150)	(230)

Notes: The results are catchment area (health facility) averages. Robust standard errors in parentheses. Significantly different from zero at 99 (***), 95 (**), and 90 (*) percent confidence level. Description of variables: See table 1c.

Table 1c. Average health facility and citizen characteristics, pre-treatment

Variables	Treatment group	Control group	Difference
User charges:			
Drugs	0.02	0.01	0.01
	(0.01)	(0.01)	(0.01)
General treatment	0.10	0.03	0.07*
	(0.04)	(0.01)	(0.04)
Delivery	0.24	0.20	0.04
	(0.04)	(0.05)	(0.06)
Injection	0.50	0.58	-0.08
	(0.08)	(0.06)	(0.10)
Average standardized pre-treatment effects:			
Utilization			0.11
			(0.77)
Utilization pattern			0.09
			(0.22)
Quality measures			0.16
			(0.54)
Catchment area statistics			0.11
			(0.66)
Health facility characteristics			0.41
			(0.29)
Citizen perceptions			0.42
			(0.63)
Supply of drugs			0.73
			(0.83)
User charges			0.65
			(0.63)

Notes: The results are catchment area (health facility) averages. Robust standard errors in parentheses. Significantly different from zero at 99 (***), 95 (**), and 90 (*) percent confidence level. Description of variables: Utilization variables are the average number of patients visiting the health facility per month. Utilization pattern is the citizens' use of different service providers in case of illness (reported in percentages). Waiting time is calculated as the difference between the time the citizen left the facility and the time the citizen arrived at the facility minus the examination time. Equipment usage is a dummy variable indicating whether the staff used any equipment during examination. Funding at the health facility is the average funds received at the health facility per month from the district and the Health Sub-district (measured in 1000 shillings). Catchment area statistics are determined from UBOS maps and census data. Piped water, Rain tank and Borehole are dummy variables indicating the health facility's water source. Drinking water is a indicator variable (1-3) on how reliable the facility's source of drinking water is (1=very reliable). Drink safely today is a dummy variable indicating whether the health facility staff at the time of the pre-intervention survey could safely drink from the water source. Separate maternity unit is a dummy variable indicating whether the health facility has a separate maternity unit. Distance to nearest Local Council I and distance to nearest public health provider is measured in kilometres. Number of days without electricity in the last month is measured out of 31 days. A level education is advanced secondary (S5-S6) education or university preparatory education. Citizen's perceptions describe his/her experience during the last visit at the health facility and are measured on a scale from 1 to 4 where a higher value represents higher satisfaction. Citizen's information about drug deliveries is a dummy variable indicating if the citizen knows when the health facility receives drugs from the district/Health Sub-district. Supply of drug deliveries per month is measured as the average number of tablets received at the health facility per month from the district/Health Sub-district. User charges are a dummy variable indicating if the household had to pay for the service provided at the health facility. Average standardized pre-treatment effects are calculated by estimating (5) on each family of outcomes using pre-intervention data.

Table 2: Program impact on monitoring tools at the health facility

Dependent variable	Suggestion box	Numbered waiting cards	Poster informing of free services	Poster on patients' rights	Average standardized effect	At least one monitoring tool	Discuss health facility in LC meetings
Specification	(i)	(ii)	(iii)	(iv)	(v)	(vi)	(vii)
Program impact	0.38***	0.20**	0.19**	0.12	2.52***	0.56***	0.13***
	(0.10)	(0.10)	(0.09)	(0.10)	(0.48)	(0.11)	(0.02)
Mean control group	0	0.04	0.12	0.12	-	0.16	0.33
District fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	50	50	50	50	50	50	3119
R2	0.35	0.30	0.47	0.26	-	0.70	0.11

Notes: Robust standard errors in parentheses. Disturbance terms are clustered by catchment areas in regression (vii). *** [**] (*) denote significance at the 1 [5] (10) percent level. Dependent variables in specifications (i)-(vi) are based on data collected through visual checks by the enumerators: (i) Dummy variable indicating if the health facility has a suggestion box for complaints and recommendations; (ii) Dummy variable indicating if the health facility has numbered waiting cards for its patients; (iii) Dummy variable indicating if the health facility has a poster informing about free health services; (iv) Dummy variable indicating if the health facility has a poster on patients' rights and obligations; (v) Average standardized effect of the monitoring variables in columns (i)-(iv) with robust standard errors derived from equation (5); (vi) Dummy variable indicating if the health facility has a least one of the "monitoring tools" (suggestion boxes, numbered waiting cards, posters on free-services), (vii) Dummy variable indicating if the household discussed the functioning of the health facility at a Local council meeting during the past year.

Table 3: Difference-in-difference estimates of the program impact on treatment practices

Dependent variable Specification	Equipment usage (i)	Waiting time (ii)
Program impact (Treatment*2005)	0.08** (0.03)	-14.0* (7.7)
2005	-0.07*** (0.02)	-9.6* (5.3)
Mean control group 2005	0.41	133
Facility fixed effects	Yes	Yes
Observations	5280	5148
R2	0.15	0.12

Notes: Robust standard errors clustered by catchment areas in parentheses. Significantly different from zero at 99 (***), 95 (**), and 90 (*) percent confidence level.
Specification: (i) Dummy variable indicated whether the staff used any equipment during examination when the citizen visited the health facility; (ii) Difference between the time the citizen left the facility and the time the citizen arrived at the facility minus the examination time.

Table 4: Program impact on management and citizens' information

Dependent variable Specification	Absence rate (i)	Staff present (ii)	Health information (iii)	Importance of family planning (iv)
Program impact	-0.14** (0.06)	0.74** (0.22)	0.07*** (0.02)	0.06*** (0.02)
District fixed effects	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes
Observations	46	46	4996	4996
R2	0.63	0.50	0.10	0.11

Notes: Robust standard errors in parentheses. Disturbance terms are clustered by catchment areas (specification (iii)-(iv)). *** [**] (*) denote significance at the 1 [5] (10) percent level.
Specification: (i) Absence rate is the ratio of workers not physically present at the time of the post-intervention survey to the number of workers employed pre-intervention (see text for details); (ii) Staff present is the number of workers (verified by the enumerators) to be present at the time of the (surprise) post-intervention survey; (iii) Dummy variable indicating if the household has received information about the importance of visiting the health facility and the danger of self-treatment, (iv) Dummy variable indicating if the household has received information about family planning. Controls are measured pre-treatment and include number of villages in catchment area, number of days without electricity in the last month, dummy for separate maternity unit, distance to nearest public health provider, number of staff with less than advanced A-level education, dummy indicating if drinking water is safe at the time of the pre-intervention survey, and average monthly supply of Quinine.

Table 5a: Program impact on immunization of newborns

Dependent variable	Polio	DPT	BCG	Vitamin A	Polio	DPT	BCG	Vitamin A	Average standardized effect	Average standardized effect
Specification	(i)	(ii)	(iii)	(iv)	(v)	(vi)	(vii)	(viii)	(ix)	(x)
Program impact	0.15*	0.04	0.22***	0.10**	0.10	0.03	0.22**	0.11*	1.78***	1.27**
	(0.08)	(0.06)	(0.08)	(0.05)	(0.07)	(0.06)	(0.09)	(0.06)	(0.67)	(0.60)
Constant	0.36***	0.21***	0.48***	0.08**					-	-
	(0.06)	(0.05)	(0.06)	(0.03)						
District fixed effects	No	No	No	No	Yes	Yes	Yes	Yes	No	Yes
Controls	No	No	No	No	Yes	Yes	Yes	Yes	No	Yes
Observations	173	173	173	173	173	173	173	173	173	173

Notes: Robust standard clustered by catchment areas in parentheses. *** [**] (*) denote significance at the 1 [5] (10) percent level. Sample includes children under 3 months. Point estimates, standard errors, and average standardized effect are derived from equation (5). Dependent variables in specifications (i)-(viii) are dummy variables indicating if the child has received at least one dose of DPT, BCG, polio, and Vitamin A supplement. Specification (ix)-(x): average standardized effect of DPT, BCG, polio, and Vitamin A. Control variables: see notes to table 4.

Table 5b: Program impact on immunization of children less than 1-year old

Dependent variable	Polio	DPT	BCG	Vitamin A	Polio	DPT	BCG	Vitamin A	Average standardized effect	Average standardized effect
Specification	(i)	(ii)	(iii)	(iv)	(v)	(vi)	(vii)	(viii)	(ix)	(x)
Program impact	0.05	0.01	0.08**	0.05	0.05	0.01	0.06**	0.05	1.08	1.34*
	(0.03)	(0.04)	(0.04)	(0.14)	(0.03)	(0.03)	(0.03)	(0.07)	(0.72)	(0.73)
	0.40***	0.36***	0.76***	0.58**					-	-
	(0.03)	(0.03)	(0.03)	(0.10)						
District fixed effects	No	No	No	No	Yes	Yes	Yes	Yes	No	Yes
Controls	No	No	No	No	Yes	Yes	Yes	Yes	No	Yes
Observations	961	961	961	961	961	961	961	961	961	961

Notes: See notes to table 5c. Sample includes children 0-12 months.

Table 5c: Program impact on immunization of 1-year olds

Dependent variable	Measles	Polio	DPT	BCG	Vitamin A	Measles	Polio	DPT	BCG	Vitamin A	Average standardized effect	Average standardized effect
Specification	(i)	(ii)	(iii)	(iv)	(v)	(vi)	(vii)	(viii)	(ix)	(x)	(xi)	(xii)
Program impact	0.05*	0.05*	0.08**	0.06	0.12	0.03	0.06**	0.06***	0.00	0.03	1.21*	1.25*
	(0.03)	(0.03)	(0.04)	(0.04)	(0.16)	(0.02)	(0.03)	(0.02)	(0.02)	(0.08)	(0.73)	(0.74)
	0.79***	0.79***	0.75***	0.92***	1.13***						-	-
	(0.02)	(0.03)	(0.03)	(0.02)	(0.12)							
District fixed effects	No	No	No	No	No	Yes	Yes	Yes	Yes	Yes	No	Yes
Controls	No	No	No	No	No	Yes	Yes	Yes	Yes	Yes	No	Yes
Observations	979	979	979	979	979	979	979	979	979	979	979	979

Notes: Robust standard clustered by catchment areas in parentheses. *** [**] (*) denote significance at the 1 [5] (10) percent level. Sample includes children 13-24 months. Point estimates, standard errors, and average standardized effect are derived from equation (5). Dependent variables in specifications (i)-(x) are dummy variables indicating if the child has received the required dose(s) of measles, DPT, BCG, polio, and number of doses of Vitamin A supplement. Specification (xi)-(xii): average standardized effect of measles DPT, BCG, polio, and Vitamin A. Control variables: see notes to table 4.

Table 5d: Program impact on immunization of 2-year olds

Dependent variable	Measles	Polio	DPT	BCG	Vitamin A	Measles	Polio	DPT	BCG	Vitamin A	Average standardized effect	Average standardized effect
Specification	(i)	(ii)	(iii)	(iv)	(v)	(vi)	(vii)	(viii)	(ix)	(x)	(xi)	(xii)
Program impact	0.01	0.03	0.03	-0.00	-0.05	0.01	0.04**	0.06***	-0.00	-0.08	0.34	1.04
	(0.02)	(0.02)	(0.04)	(0.01)	(0.16)	(0.02)	(0.02)	(0.02)	(0.01)	(0.09)	(0.78)	(0.71)
	0.88***	0.85***	0.80***	0.94***	1.32***						-	-
	(0.02)	(0.02)	(0.02)	(0.01)	(0.12)							
District fixed effects	No	No	No	No	No	Yes	Yes	Yes	Yes	Yes	No	Yes
Controls	No	No	No	No	No	Yes	Yes	Yes	Yes	Yes	No	Yes
Observations	995	995	995	995	995	995	995	995	995	995	995	995

Notes: See notes to table 5c. Sample includes children 25-36 months.

Table 5e: Program impact on immunization of 3-year olds

Dependent variable	Measles	Polio	DPT	BCG	Vitamin A	Measles	Polio	DPT	BCG	Vitamin A	Average standardized effect	Average standardized effect
Specification	(i)	(ii)	(iii)	(iv)	(v)	(vi)	(vii)	(viii)	(ix)	(x)	(xi)	(xii)
Program impact	0.03**	0.05**	0.10***	0.02	-0.05	0.04***	0.04***	0.10***	0.01	-0.12	1.79**	1.92**
	(0.01)	(0.02)	(0.03)	(0.01)	(0.19)	(0.01)	(0.02)	(0.02)	(0.01)	(0.08)	(0.79)	(0.76)
Constant	0.91***	0.87***	0.80***	0.93***							-	-
	(0.01)	(0.02)	(0.03)	(0.01)								
District fixed effects	No	No	No	No	No	Yes	Yes	Yes	Yes	Yes	No	Yes
Controls	No	No	No	No	No	Yes	Yes	Yes	Yes	Yes	No	Yes
Observations	1163	1163	1163	1163	1163	1163	1163	1163	1163	1163	1163	1163

Notes: See notes to table 5c. Sample includes children 37-48 months.

Table 5f: Program impact on immunization of 4-year olds

Dependent variable	Measles	Polio	DPT	BCG	Vitamin A	Measles	Polio	DPT	BCG	Vitamin A	Average standardized effect	Average standardized effect
Specification	(i)	(ii)	(iii)	(iv)	(v)	(vi)	(vii)	(viii)	(ix)	(x)	(xi)	(xii)
Program impact	0.01	0.03	0.03	0.01	0.03	0.00	0.03	0.04	0.02	-0.08	0.60	0.75
	(0.02)	(0.03)	(0.02)	(0.02)	(0.19)	(0.02)	(0.02)	(0.03)	(0.02)	(0.16)	(0.80)	(0.82)
Constant	0.94***	0.89***	0.88***	0.95***							-	-
	(0.02)	(0.02)	(0.03)	(0.02)								
District fixed effects	No	No	No	No	No	Yes	Yes	Yes	Yes	Yes	No	Yes
Controls	No	No	No	No	No	Yes	Yes	Yes	Yes	Yes	No	Yes
Observations	542	542	542	542	542	542	542	542	542	542	542	542

Notes: See notes to table 5c. Sample includes children 49-60 months.

Table 6: Program impact on utilization/coverage

Dependent Variable	Out-patients			Delivery			Antenatal		Family planning		Average standardized effect	
Specification	(i)	(ii)	(iii)	(iv)	(v)	(vi)	(vii)	(viii)	(ix)	(x)	(xi)	(xii)
Program impact	107.4*	128.9**	146.5***	6.3*	4.8**	3.5*	16.1	11.1	5.5	3.8	1.44**	1.68***
	(64.4)	(60.1)	(47.8)	(3.3)	(2.0)	(2.0)	(16.0)	(11.9)	(4.8)	(3.1)	(0.70)	(0.60)
Pre-utilization			0.48***			1.0***						
			(0.07)			(0.09)						
Constant	660.8***			9.2***			78.9***		15.2***			
	(34.6)			(1.6)			(11.8)		(3.5)			
District fixed effects	No	Yes	No	No	Yes	No	No	Yes	No	Yes	No	Yes
Controls	No	Yes	No	No	Yes	No	No	Yes	No	Yes	No	Yes
Observations	50	50	50	50	50	50	50	50	50	50	50	50
R2	0.05	0.42	0.51	0.07	0.64	0.61	0.02	0.66	0.03	0.46	-	-

Notes: Robust standard errors in parentheses. *** [**] (*) denote significance at the 1 [5] (10) percent level. Point estimates, standard errors, and average standardized effects are derived from equation (5). Dependent variables are monthly averages of patients seeking care. Control variables: see notes to table 4.

Table 7: Difference-in-differences estimates of the program impact on citizens' health seeking behavior

Dependent variables	Project facility	NGO	Private-for-profit	Traditional healer	Self-treatment	Traditional healer & self treatment	Other-government Facility	Other
Specification	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Program impact	0.029*	-0.003	0.026	-0.013*	-0.029	-0.042**	-0.000	-0.010
(Treatment*2005)	(0.016)	(0.006)	(0.020)	(0.007)	(0.020)	(0.019)	(0.011)	(0.018)
2005	-0.07***	0.007	-0.02	-0.002	0.034**	0.033**	-0.010	0.047***
	(0.011)	(0.005)	(0.014)	(0.005)	(0.012)	(0.012)	(0.007)	(0.013)
Mean control group 2005	0.23	0.03	0.22	0.03	0.34	0.37	0.12	0.06
Facility fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	9319	9319	9319	9319	9319	9319	9319	9319
R2	0.08	0.09	0.05	0.03	0.10	0.09	0.07	0.08

Notes: Robust standard clustered by catchment areas in parentheses. *** [**] (*) denote significance at the 1 [5] (10) percent level. Dependent variables are citizens' use of service providers in case of illness (reported in percentages).

Table 8: Program impact on health outcomes: Under-five child deaths

Dependent variable	Child death (children < 5 year)		Under-5 mortality rate	
Specification	(1)	(2)	(3)	(4)
Program impact	-0.016*	-0.016*	-48.0*	-53.2**
	(0.01)	(0.009)	(24.2)	(25.3)
Constant	0.049***		144.9***	
	(0.006)		(16.9)	
Controls	No	Yes	No	Yes
District fixed effects	No	Yes	No	Yes
Observations	2922	2922	50	50
R2	0.002	0.01	0.08	0.34

Notes: Robust standard errors in parentheses (iii)-(iv), clustered by catchment area (i)-(ii). *** [**] (*) denote significance at the 1 [5] (10) percent level. Dependent variable in specification (i)-(ii) is a dummy variable indicating whether any children under five in the household have died during the last year and estimated under-5 mortality rate in the community expressed per 1,000 live births (specification (iii)-(iv)). Control variables: see notes to table 4.

Table 9: Program impact on health outcomes: Child weight of infants

Dependent variable	Weight-for-age z-scores				
	1-18 months			19-36 months	
Specification	(1)	(2)	(3)	(4)	(5)
Program impact	0.16*	0.17**	0.14**	0.012	0.09
	(0.09)	(0.08)	(0.07)	(0.09)	(0.06)
Child age (log)			-1.27***		0.06
			(0.07)		(0.17)
Female			0.27***		0.08
			(0.09)		(0.07)
Constant	-0.64***	-0.71***		-0.95***	
	(0.07)	(0.06)		(0.08)	
Controls	No	No	Yes	No	Yes
District fixed effects	No	No	Yes	No	Yes
Observations	1167	1135	1135	1300	1300
R2	0.002	0.004	0.22	0.00	0.04

Notes: Robust standard clustered by catchment area in parentheses. *** [**] (*) denote significance at the 1 [5] (10) percent level. Dependent variable is weight-for-age z-scores. Specification: (i) all children under 18 months; (ii)-(iii) all children under 18 months excluding observations with recorded weight above the 90th percentile in the growth chart reported in Cortinovis et al (1997); (iv)-(v) all children between 18 and 36 months. Control variables: see notes to table 4.

Table 10: Robustness test: The effect of treatment on utilization in the control group

Dependent variable	Out-patients	Delivery	Family planning	Antenatal care
Specification	(1)	(2)	(3)	(4)
Distance to closest project facility in 2005	-1.13	-0.10	0.07	-0.56
	(2.11)	(0.07)	(0.22)	(0.52)
Constant	696***	12.4***	13.0*	96.0***
	(66)	(2.7)	(7.0)	(18.1)
District fixed effects	No	No	No	No
Observations	25	25	25	25
R2	0.02	0.06	0.01	0.03

Notes: Robust standard errors in parentheses. Significantly different from zero at 99 (***), 95 (**), and 90 (*) percent confidence level. See notes to table 6.

Table 11: Robustness test: Difference-in-difference estimates of drugs supply and funding received

Dependent variable	Erythromycin	Chlorquine	Septrine	Quinine	Mebendazole	Funding
Specification	(i)	(ii)	(iii)	(iv)	(v)	(vi)
Program impact	151.8	-176	-2.9	-237	114	-248
(Treatment*2005)	(127)	(688)	(682)	(154)	(533)	(1224)
2005	-145**	-531	-457	-30	984**	1261
	(65)	(415)	(415)	(101)	(412)	(965)
Mean control group 2005	257	2384	1973	305	2483	4471
District fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Observations	96	100	100	99	100	94
R2	0.73	0.55	0.60	0.57	0.66	0.72

Notes: Robust standard errors in parentheses. Significantly different from zero at 99 (***), 95 (**), and 90 (*) percent confidence level. Dependent variables are average number of tablets received at the health facility per month from the district and Health Sub-district during the last year (specification (i)-(v)); average amount of public health care funds received per month from the district and Health Sub-district during the last year (measured in 1000 Uganda shillings).

Table 12: Robustness test: Program impact on monitoring of upper-level authorities and dismissals and transfers of staff

Dependent variable	Sub-country official	Parish official	Dismissals	Transferred	Left	New health workers
Specification	(i)	(ii)	(iii)	(iv)	(v)	(vi)
Program impact	0.12	0.06	-0.05	-0.08	0.08	0.52
(Treatment*2005)	(0.12)	(0.11)	(0.07)	(0.23)	(0.15)	(0.35)
District fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Observations	50	50	50	50	50	50
R2	0.73	0.55	0.60	0.57	0.66	0.25

Notes: Robust standard errors in parentheses. Significantly different from zero at 99 (***), 95 (**), and 90 (*) percent confidence level. Dependent variables are (i) dummy indicating if the facility has received a monitoring/support visit from any Sub-county officials in 2005; (ii) dummy indicating if the facility has received a monitoring/support visit from any Parish officials in 2005; (iii) number of staff that has been dismissed in 2005; (iv) number of staff that has been transferred from the facility in 2005; (v) number of staff that voluntarily left the facility in 2005; (vi) number of new workers.

Figure 4: Growth charts

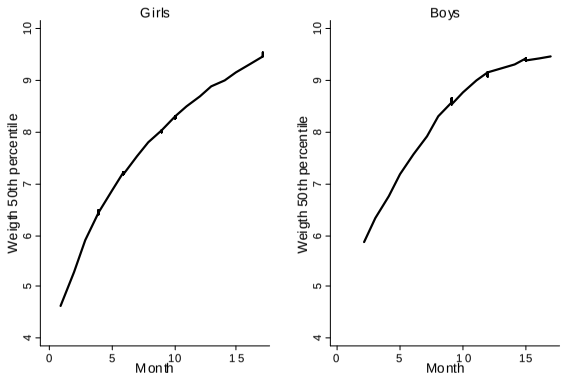
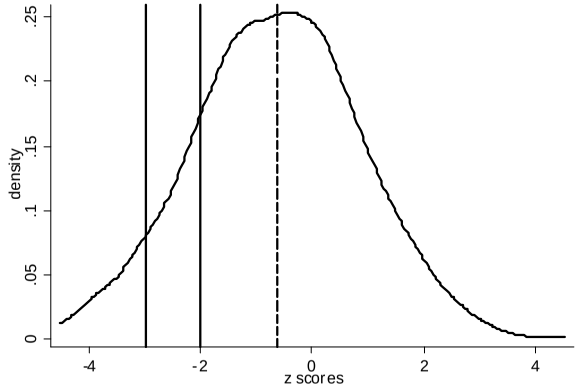
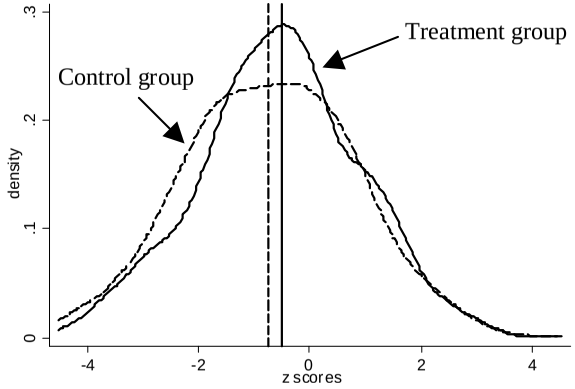


Figure 5: Distribution of weight-for-age z scores for infants (1-18 months)



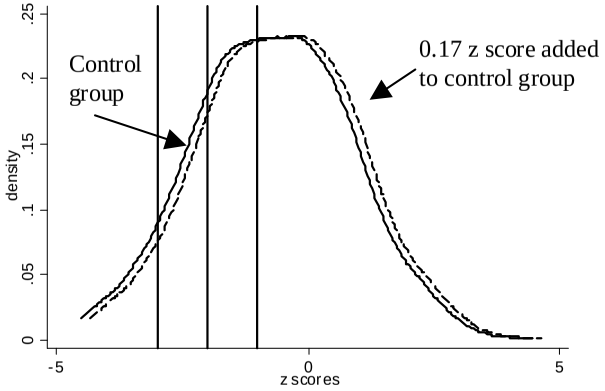
Notes: Vertical solid lines denote -3 and -2 z scores, dashed line denotes mean z-score

Figure 6: Distributions of z scores for treatment and control groups



Notes: Vertical solid line denotes mean in treatment group, dashed line denotes mean in control group.

Figure 7: Treatment effect



Notes: Vertical solid lines denote -3, -2, -1 z scores.