The Weakness of Bottom-Up Accountability: Experimental Evidence from the Ugandan Health Sector

Pia Raffler [*] Daniel N. Posner [†] Doug Parkerson [‡]

This version: March 26, 2019

Abstract

We evaluate the impact of a large-scale information and mobilization intervention designed to improve health service delivery in rural Uganda by increasing citizens' ability to monitor and apply bottom-up pressure on underperforming health workers. Modeled closely on the landmark "Power to the People" study (Björkman and Svensson, 2009), the intervention was undertaken in 376 health centers in 16 districts and involved a three-wave panel of more than 14,000 households. We find that while the intervention had a modest positive impact on treatment quality and patient satisfaction, it had no effect on utilization rates or health outcomes (including child mortality). We also find no evidence that the channel through which the intervention affected treatment quality was citizen monitoring. The results hold in a wide set of pre-specified subgroups and also when, via a factorial design, we break down the complex intervention into its two most important components. Our findings cast doubt on the power of information to foster community monitoring or to generate improvements in health outcomes, at least in the short term.[1]

JEL codes: I12, O15, C93, D73, H75

Keywords: accountability, information, child mortality, health care, Uganda, field experiment

[*]Department of Government, Harvard University, praffler@gov.harvard.edu.

[†]Department of Political Science, University of California, Los Angeles, dposner@polisci.ucla.edu.

[‡]Innovations for Poverty Action, dparkerson@poverty-action.org.

[1]Authors listed in reverse alphabetical order. We thank the Ministry of Health in Uganda, in particular Dr. Anthony Mbonye; IPA Uganda, in particular Frédéric Cochinard, Martin Atyera, Joshua Bwiira, Paola Elice, Afke Jager, Kyle Holloway, Douglas Kaziiro, Steven Kizza, Ezra Rwakazooba, Laura Schmucker, and Alex Tusiime for excellent research assistance and Damien Kirchhoffer, Dickson Malunda, and Daniele Ressler for their dedicated project management. We thank our partners at GOAL Uganda, in particular Elizabeth Allen, Angela Bailey, Niamh Barry, and Fiona Mitchell; our implementing partners, Coalition for Health Promotion and Social Development, Kabarole Research and Resource Centre, and Multi-Community Based Development Initiative; and the Department for International Development for funding. Helpful comments were provided by Ala' Alrababa'h, Graeme Blair, Kate Casey, Darin Christensen, Pascaline Dupas, Daniel de Kadt, Don Green, Saad Gulzar, Johannes Haushofer, Chad Hazlett, Stuti Khemani, Stephen Kosack, Will Marble, Ted Miguel, Rachel Myrick, Gareth Nellis, Melina Platas, Francesco Trebbi, participants at the 17th and 23rd EGAP meetings, and at seminars at Columbia, Emory, Rochester, Stanford, UC Berkeley, UCLA, Vancouver School of Economics, and World Bank. The pre-analysis plan was registered at EGAP (ID 20160611AA). Human subjects approval was received under IPA IRB #2127. A working paper version of this paper was submitted for pre-publication re-analysis to the Abdul Latif Jameel Poverty Action Lab (J-PAL), where a code replication exercise was undertaken. We thank Georgiy Syunyaev and Isabelle Cohen for conducting this re-analysis.

1 Introduction

Poor service delivery is a major problem in developing countries, particularly in primary health care. In the world's poorest countries, staff at rural government-run clinics are often absent, adherence to clinical guidelines is weak, shortages of basic drugs are common, and services such as family planning and antenatal care are underprovided.[2] In part for these reasons, utilization rates at government clinics are low. Many children fail to receive essential vaccinations. Stunting and anemia are common. Under-five mortality rates, although declining, are still more than ten times higher than in developed countries (UNICEF, 2017). Improving the delivery of primary health services is therefore a central development priority (World Bank, 2004).

In recent years, development funders and practitioners have embraced a potentially promising approach to the problem of poor service provision: the bottom-up monitoring of service providers by community members (Mansuri and Rao, 2013; Kosack and Fung, 2014; Molina et al., 2016). Rooted in the logic of the principal-agent problem, the idea is that providing citizens with information about service delivery shortfalls, along with information allowing them to compare local outcomes with national standards and with outcomes in other communities, will put them in a position to monitor and apply pressure on underperforming service providers. The presumed causal arrows run from information to citizen pressure to improved provider behavior to improvements in health outcomes. A major attraction of this approach is that it leverages the growing space for political engagement in many developing countries while directly addressing the lack of effort and corruption of service providers that is viewed as one of the major sources of poor service delivery in such settings (World Bank, 2016). The approach has the additional appeal of attacking the problem without requiring expensive inputs such as additional staff, training, or new equipment.

The attractiveness of this bottom-up, information-focused, community monitoring strategy was validated by the findings of a landmark randomized study published in 2009 by Martina Björkman and Jakob Svensson (Björkman and Svensson, 2009). The "Power to the People" (P2P) study sought to improve local health care provision in rural Uganda by providing community members and local health care providers with information about the quality of health services being provided at the local government-run health center (HC) and then bringing the community members and health center staff together to discuss how they might collaborate to improve health outcomes in the community. The P2P intervention generated striking results: infant weights increased in treatment communities; under-5 mortality declined by 33 percent; immunization rates rose; waiting times at clinics fell; staff absenteeism dropped; utilization increased; and communities became

[2]For country-specific details, see the reports generated as part of the World Bank's Service Delivery Indicators project, http://datatopics.worldbank.org/sdi/.

more engaged and monitored health providers more extensively.[3] Given both these large effects and the appeal of the approach it was testing, the P2P study received wide acclaim. It has been held up as an example of the power of information to generate accountability and of the utility of community-based monitoring as a tool for improving health outcomes in developing country settings. Hundreds of millions of dollars have been spent on programming inspired by the P2P design.

Notwithstanding the strong findings in P2P, the effectiveness of similar programs as tools for improving frontline service provision has received only mixed support in other research. Olken (2007), Banerjee et al. (2010), and Keefer and Khemani (2014) all report weak effects of interventions designed to generate behavioral change by frontline service providers through information provision and bottom-up grassroots monitoring. Pandey, Goyal and Sundararaman (2009), Barr et al. (2012), Pradhan et al. (2009), Andrabi, Das and Khwaja (2017), Fiala and Premand (2018), and Banerjee et al. (2018), meanwhile, find more promising effects. In the study closest to our own, Dube, Haushofer and Siddiqi (2018) find effects of a bottom-up health intervention in Sierra Leone on utilization and child mortality, but not on service quality or other health outcomes. These mixed results, combined with the limited power of the original P2P study—the intervention included just twenty-five treated health centers—have raised questions about how certain we can be about the power of information provision and community monitoring to improve service delivery.

In this paper, we report the results of Accountability Can Transform (ACT) Health, a large-scale intervention designed to improve health service delivery in rural Uganda.[4] Modeled on P2P, the objective of ACT Health was to learn more about the strengths, limitations, and operation of the causal pathway that P2P popularized. ACT Health randomized the delivery of information about patient rights and responsibilities, utilization patterns and health outcomes at the local health center, worked with health center staff and community members to develop action plans in light of that information, and organized meetings between members of the community and health center staff to generate a joint social contract to guide both actors' future behavior and interactions. The intervention was implemented in 376 health centers and their associated catchment areas in 16 districts. The study involved the collection of three waves of panel data on utilization rates, treatment quality, patient satisfaction, and health outcomes (including child mortality) at both the health center and household (N=14,609) levels. To capture the channels through which the intervention was hypothesized to effect change, we collected data on a broad array of intermediate outcomes

[3] In a follow-up paper, Björkman Nyqvist, de Walque and Svensson (2017) show that these positive effects persist four years after the initial intervention.

[4] ACT Health was implemented by GOAL Uganda with funding from the UK's Department for International Development. The evaluation described in this paper was undertaken by Innovations for Poverty Action under the direction of the study authors. ACT Health included additional advocacy components in a second, follow-on phase, which we are not evaluating here.

as well as on health center, community and household characteristics that might be sources of differential treatment effects in particular subgroups of health centers and catchment areas. We also implemented a factorial design to gain a deeper understanding of the mechanisms at work. Given the project's scope and the comprehensiveness of the data we collected, our study provides a particularly high-powered test of the potential impact of information provision and community monitoring on primary health outcomes in a developing country setting.

We find no statistically significant effects on utilization rates or health outcomes (including child mortality), although we do find positive (albeit substantively small) impacts on treatment quality and patient satisfaction. These results are reinforced when we examine sub-populations of health centers, communities and individuals: we find persistently null effects on utilization and health outcomes across nearly all subgroups and we find larger impacts on treatment quality in subgroups in which we might have expected to find stronger effects. The null results on health outcomes and child mortality also hold in both of the treatment arms we investigate via the factorial design: 1) the provision of information and the mobilization of health center staff and community members in light of that information, and 2) the holding of interface meetings in which health center staff and citizens can confront one another and work together to develop a plan of action to improve health outcomes. We also find little evidence that the intervention caused citizens to increase their monitoring or sanctioning of health care workers, although we find suggestive evidence that the presence of sub-county officials during the programming boosted the impact of the intervention on treatment quality. Consistent with the conclusions in World Bank (2016), this suggests that top-down monitoring by government officials may be a more powerful tool for changing health workers' behavior than bottom-up monitoring by citizens. Taken together, our findings cast doubt on the ability of information to generate community monitoring or improvements in bottom-line health outcomes.

These findings contrast sharply with those reported in P2P. As we discuss in greater detail in Section 7, a plausible explanation for these differences lies in the very different baseline conditions in the two studies.[5] While child mortality rates at the time of P2P were 117 per 1,000 live births, they had decreased to 59 per 1,000 by the time of the ACT Health baseline—much closer to the current median rate in Sub-Saharan Africa.[6] Indeed, as of 2017 only five countries in Sub-Saharan Africa were within one half a standard deviation of the child mortality rates in Uganda at the time of P2P.[7] Our findings may therefore be particularly relevant for our understanding of how to

[5] As we report in Section 7, we find significant child mortality results (although no treatment impacts on utilization, treatment quality, or other health outcomes) in the sub-sample of health centers whose baseline child mortality rates are within one standard deviation of those reported at baseline in P2P.

[6] Data from World Development Indicators.

[7] One of those countries is Sierra Leone, which may account for why Dube, Haushofer and Siddiqi (2018) find significant treatment effects in a P2P-inspired intervention designed to improve primary health outcomes through

improve health outcomes in developing countries today.

2 Health Service Delivery in Rural Uganda

Public health services in Uganda are provided in a hierarchical system with national referral hospitals at the national level, regional referral hospitals at the regional level, general hospitals at the district level, and smaller scale health centers at the sub-county and parish levels—the former termed HC3s; the latter, HC2s. Our study focuses on health care delivery at the HC3 and HC2 levels, the lowest levels of the public health system. HC3s, which are staffed by a trained medical worker and one or more nurses and lab technicians, provide preventative and out-patient care and have laboratory services to undertake basic tests.[8] They also generally have maternity wards and offer prenatal and antenatal services. HC2s provide outpatient services and antenatal care. They are run by a nurse, sometimes working with a midwife and a nursing assistant. Both types of units are supported by Village Health Teams (VHTs) comprised of volunteer community health workers who undertake health education outreach, provide simple curative services, and refer patients to higher level health centers for treatment of more complicated conditions. Generally speaking, patients seek care at the facility closest to their home and are then referred on to higher-level facilities as the nature of their medical condition requires.

Government-run health facilities operate alongside a growing number of private for-profit and not-for-profit (often religious) health providers, as well as traditional practitioners. In our sample at baseline, 40 percent of households that reported having a health condition requiring treatment during the past year sought care at a government-run health center, whereas 18 percent sought care at a private clinic. Thirty-three percent self-treated.[9] Among the reasons cited for not visiting the government-run health center were lack of drugs, long waiting times, poor quality of services, and poor staff attitude. Just 60 percent of households that sought care at the government-run health center said that the staff clearly explained their diagnosis and only 46 percent judged the services they received to be of "very high" or "somewhat high" quality. At baseline, only 27 percent of health center staff were present during an unannounced visit.[10]

Factors both within and outside the health workers' control contribute to these outcomes. Un-

bottom-up citizen pressure.

[8] These are the government standards. HC3s frequently do not have the full set of staff or provide the full set of services that government standards specify.

[9] Seven percent sought care from a member of the VHT and two percent sought care from a traditional healer.

[10] Although our sample was not drawn to be nationally representative, these findings are consistent with data collected on utilization and satisfaction with health outcomes in Uganda more broadly (Rutaremwa et al., 2015; Uganda Bureau of Statistics, 2017).

derstaffing, low and irregular pay, shortages of necessary medical supplies, and limited oversight by higher-level health officials are major problems (Uganda Ministry of Health, 2017). They lead to low morale, absenteeism, and poor treatment quality, which in turn generate poor health outcomes and reduce incentives for citizens to utilize the government-run health facilities.

3 Intervention

The ACT Health intervention was implemented by a consortium of civil society organizations, coordinated by GOAL Uganda.[11] Like most other interventions in the "transparency and accountability" space (Kosack and Fung, 2014), the theory of change underlying ACT Health was that service delivery could be improved by empowering community members to demand high quality services, monitor service providers, and hold them accountable for poor performance. This is the "short route" of accountability popularized in the World Bank's influential 2004 World Development Report (World Bank, 2004).

The intervention consisted of three components, closely modeled on P2P.[12]

Information. The research team used data collected in the baseline health center and household surveys to create citizen report cards (CRCs) providing health center-specific information about citizens' knowledge of their rights and responsibilities, utilization of the various services offered at the health center, citizens' perceptions of the quality of these services, and overall satisfaction with the health care they received. For most outcomes, the health center-specific data was presented alongside district averages to provide a benchmark of relative performance. The CRCs were shared with both health care providers and community members. Information was presented with the help of visual props designed by local artists to ensure comprehension among illiterate participants.

Mobilization. Trained facilitators worked with local leaders and VHT members to organize community meetings at which the CRC results were presented and discussed. An action plan was developed to identify specific steps that could be taken by community members to improve health service delivery. Significant efforts were made to ensure that the meetings included representatives from all major social groups in the community.[13] Parallel meetings were also held separately with

[11]The project was approved by the Internal Review Boards at IPA (Protocol ID: 0497) and at the Uganda National Council for Science and Technology (UNCST) (Protocol ID: ARC157). Approval for the project was also received from UNCST itself (Protocol ID: SS3559) and the Office of the President, Uganda. Participation in the study was voluntary and all respondents needed to give their informed consent in order to participate. Respondents did not receive any compensation for their participation.

[12]A summary of these components, taken from the training manual developed by GOAL, is provided in Appendix H. The deviations from P2P in program design and implementation are summarized in Appendix G.

[13]The meetings included an average of 100 attendees. Further details about the meeting participants, as well as the

health center staff at which the CRC results were discussed and an action plan was formulated describing steps that the staff could take to improve health outcomes.

Interface. Facilitators brought the health center staff together with representatives of the community to discuss their respective action plans and how they might work together to improve the quality of health care in the community.[14] The output of the interface meeting was a social contract between the citizens and health care workers laying out specific steps that each could take to contribute to improvements in health outcomes.

Implementing teams spent several days in each catchment area to organize the community and health center dialogues and the interface meetings, and they returned every six months (for a total of three follow-up visits before endline data collection) to meet with community members and health center staff to check on the progress that had been made toward the commitments stipulated in the social contract. A time line of the intervention is provided in Figure 1. Examples of a CRC, community and health center action plans, and a joint social contract are included in Appendix H.

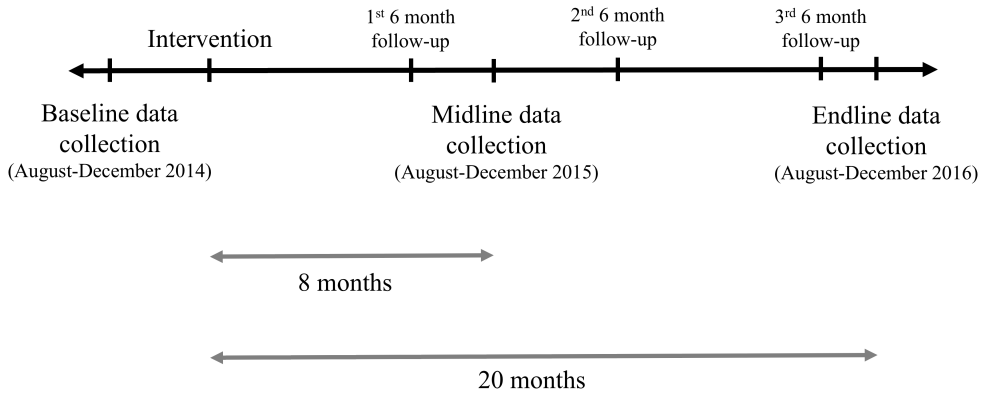


Figure 1: Time line of the intervention

The logic of the bottom-up accountability approach suggests that the information, mobilization and interface components should generate improvements in service delivery and health outcomes via three mutually-reinforcing mechanisms. First, the receipt of information by both community members and health providers, via the CRC, should increase knowledge about issues related to health care, such as patients' right and responsibilities, the services that are supposed to be offered at the local health center, and how the health outcomes and treatment practices at the local health center compare with those of other health facilities and with national standards. This information

worksheet used to guide the implementing teams' mobilization efforts, are provided in Appendix H.

[14] On average, 50 community members and four health center staff members participated in the interface meetings. Further details are provided in Appendix H.

should put citizens in a stronger position to evaluate whether their own health center is performing adequately and create common knowledge among community members and health center staff about the health center's performance.

Second, the holding of meetings to mobilize community members and the development of action plans in light of the information provided in the CRC should allow citizens to identify concrete actions that they might take to improve health outcomes. The meetings may also generate efficacy among community members, foster a sense of responsibility for monitoring health workers to make sure they provide high quality services, and help overcome free riding problems within the community—all of which may be critical for generating bottom-up pressure by citizens and behavioral changes by health center staff (Barr et al., 2012; Lieberman, Posner and Tsai, 2014).

Third, the interface meetings should provide opportunities for citizens to confront health providers directly and apply social sanctions to those revealed by the CRC to be underperforming. Alternatively, by providing a space for community members and health providers to discuss the problems and constraints they each face, the drafting of the joint social contract may generate improvements in the relationship between community members and health providers, which may in turn have positive downstream effects on utilization, service delivery, and health outcomes.

All of these aspects of the intervention should increase the ability of citizens to apply bottom-up pressure on service providers. However, there are two alternative channels, not involving citizen pressure, through which the ACT Health (and P2P) intervention(s) might also generate positive changes in service delivery and health outcomes. First, the intervention(s) might affect health outcomes through an increase in utilization rates, either by making the existence of the local government-run health center more salient or by building trust and reducing uncertainty about the monetary and non-monetary costs of seeking treatment there. If community members who are exposed to the intervention are more likely to seek professional care at the health center than to self-treat or visit traditional healers, then we would expect health outcomes to improve as a direct result of increased utilization, even in the absence of changes in community monitoring or treatment quality.

Second, the intervention may directly affect the behavior of health workers. The creation and presentation of the scorecard may make health workers feel that their behavior is being monitored, and this may cause them to put more effort into service provision. Alternatively (or in addition), hearing about the performance of their health center relative to others in the district may increase health workers' intrinsic motivation to provide better services. Thus, treatment quality—and in turn health outcomes—may improve in the absence of community monitoring.

These alternative channels—bottom-up accountability, utilization, and direct effects on health

workers—are not mutually exclusive. Our results should be interpreted with the understanding that any (or all) of these mechanisms might explain the program impacts (and non-impacts) we report below.

4 Research Design, Data, and Estimation Framework

The unit of randomization in our study is the health center and its associated catchment area. Our sample includes 376 health centers, which represent nearly the entire universe of functioning government-run HC2s and HC3s in our sixteen study districts.[15] We define the catchment area as the three villages that are closest in proximity to the health center in question (including the village in which the health center is located), as measured by the straight-line distance from the health center to the village centroid.[16] In identifying these villages, we only include villages located in the same parish (for HC2s) or sub-county (for HC3s) as the health center in question.[17]

4.1 Factorial Design

Although our primary interest is in the impact of the full ACT Health intervention, we use a factorial design to break its multifaceted treatment into two of the three main components described in Section 3. We combine the information and mobilization components into one treatment arm and cross it with the interface treatment, as depicted in Figure 2. We then randomly assign health centers and their catchment areas to one of the four treatment groups, with treatment assignment blocked by district and health center level. This design enables us to assess the effectiveness of the full ACT Health intervention by comparing units in the bottom right cell to the control group and to learn which aspects of the broader intervention are doing the work in generating the effects we find by making comparisons across all four cells.

[15]The sixteen districts are: Lira, Apac, Pader, Gulu, Lamwo, Kitgum, Agago, Katakwi, Bukedea, Manafwa, Tororo, Kabarole, Mubende, Nakaseke, Kibaale, and Bundibugyo. A map is included in Appendix G. We excluded government health centers funded by the military or prison departments because of the unique communities they serve.

[16]Catchment areas were determined using village-level shape files provided by the Uganda Bureau of Statistics (UBOS), and health center GPS coordinates collected by GOAL. To minimize overlap of catchment areas (and hence the possibility of spillovers), we excluded health centers that were less than 2.5 km apart or that shared a village among their three closest villages.

[17]If only two villages were located within a parish or sub-county, then only these two villages were included in the catchment area. In addition, if a village was split into smaller subunits (typically the village subunits would be named "A" and "B" or "1" and "2") and if field teams confirmed that this had occurred within the last 12 months (or had not been formally recognized by the appointment of a new local council), then both of these villages were included and considered as a single village.

		Interface meetings are held between community and health facility staff	
		No	**Yes**
Report card info is reported to community and health facility staff and action plans are developed	**No**	CONTROL 95 HCs (54 HC2 ; 41 HC3)	INTERFACE WITHOUT INFORMATION OR MOBILIZATION 97 HCs (60 HC2 ; 37 HC3)
	Yes	INFORMATION AND MOBILIZATION WITHOUT INTERFACE 92 HCs (55 HC2 ; 37 HC3)	FULL ACT HEALTH INTERVENTION 92 HCs (56 HC2 ; 36 HC3)

Figure 2: Factorial design

4.2 Data

Our data come from two main sources: a household survey and a health center survey. Both were collected at baseline, midline and endline, with as close as possible to 12 months separating each survey round in each health center/catchment area in order to control for seasonal effects that might influence utilization rates or health outcomes.[18] Data collection staff were completely separate from the teams that implemented the programming and had no knowledge of the treatment status of the health centers and households they visited.

Since treatment could not be administered until after the baseline data had been collected and distilled into the CRCs, the average interval between intervention and midline data collection was less than one year (8 months; SD=1.37 months). The average interval between the intervention and endline data collection was 20 months (SD=1.34 months). Given the lack of good theory to guide us on how long it should take for the treatment to generate measurable changes in actors' behavior or health outcomes (or how quickly these effects may decay), estimating program impact at two different intervals is useful. In the results presented below, we privilege the endline findings, but we report the full midline results in Appendix E.

[18]The average interval between baseline and midline surveys was 11.9 months (SD=0.3 months); the average interval between the midline and endline surveys was 12.0 months (SD=0.11 months). These intervals are balanced across treatment arms.

The health center survey consisted of three components. The first was a brief questionnaire completed at the time of initial contact with the health center in each survey round. Since this visit was unannounced, it provided an opportunity for the collection of information about staff attendance, cleanliness, wait times and other clinic characteristics before the clinic staff was able to respond to the fact that it was being evaluated. The second component was the main health center staff survey, which collected information about the variety and quality of health services provided, utilization rates, staff structure and perceptions, funding mechanisms and drug stock-outs. This survey was conducted with the most senior health center staff member, as well as randomly drawn health workers.[19] The third component involved the collection of administrative data on file at the health center, including monthly Health Management Information System (HMIS) forms and drug stock cards. Physical checks of drug stocks were conducted to verify the accuracy of these records.

The household survey was enumerated based on a baseline sampling frame of households containing at least one child under five years old or a pregnant woman, based on village household lists and consultations with the village chairperson, VHT members, Health Unit Management Committee (HUMC) members and other knowledgeable persons.[20] We randomly sampled 40 households per catchment area from this frame, with the number of households drawn from each village proportional to the number of eligible households in that village.[21]

The primary respondent for the household survey was the female head of household. The survey collected information about household members' recent experiences with the local health center (including their satisfaction with the quality of care they received), their knowledge about their rights and responsibilities, their health status, and their participation in community activities (including those directly related to monitoring the performance of their local health care providers), among other topics. All household surveys also included an anthropometric survey component in which we recorded the weight, height and middle-upper arm circumference (MUAC) of each child under the age of five in the household. The ages of the children, and their immunization status, were verified using immunization cards, if available. At endline, we also collected retrospective information on the month of birth and, if applicable, death of all children recorded at baseline and midline in order to generate more precise estimates of child mortality rates, as described in Appendix D.

[19] If the in-charge was unavailable, we interviewed the next most highly ranked (or longest serving) health center staff member. In order not to distract health workers from performing their duties, enumerators were instructed to suspend the survey when a health worker was busy and to resume when she was again available.

[20] In instances in which our informants were unsure about the ages of children in a particular household, we verified this information by visiting the household with a knowledgeable person from the village.

[21] During the baseline only, an additional short survey was administered to another 15 households in catchment areas assigned to the information and mobilization treatments (i.e., units along the bottom row in Figure 2, that received CRCs). These additional households were included to reduce noise in the measures included in the CRC and to increase the likelihood that the community would feel that the CRC represented its views and experiences.

The household surveys were conducted in ten local languages with the help of 279 field staff hired and trained by IPA Uganda. All data was collected using smart phones, with date and time stamps, GPS coordinates, and information transmitted to an encrypted server on a daily basis.[22] In all, we completed 15,295 household surveys at baseline, 14,459 at midline, and 14,609 at endline.[23] Thanks to detailed tracking protocols, we were successful in re-interviewing 95.5 percent of our study households at endline. Our analyses are therefore based on a panel of 14,609 households, each interviewed at minimum at baseline and endline, and the vast majority at three different points in time. As shown in Appendix C, the small degree of attrition we experienced is balanced across treatment arms.

4.2.1 Outcomes of Interest

We estimate the impact of the ACT Health intervention on five categories of outcomes: utilization rates, treatment quality, patient satisfaction, health outcomes, and child mortality. Child mortality is, of course, also a health outcome, but we break it out as a separate category because of its singular importance as a bottom-line measure of health system performance. For each of the first four outcome categories, we create an averaged z-score index (Kling, Liebman and Katz, 2007), constructed so that higher values imply a more positive outcome. The index can be interpreted as the average of the included measures, scaled to standard deviation units. Child mortality is calculated at the health center level via a set of indicator variables for whether each child is dead or alive in a given month.[24] The components of the five main outcome indices, along with their mean values at baseline, are presented in Table 1.

In addition to these five main outcomes, we also test for treatment effects on averaged z-score indices of seven intermediate outcomes that map onto the mechanisms discussed in Section 3: citizen knowledge, health center staff knowledge, efficacy, community responsibility, community monitoring, the relationship between health workers and the community, and health center transparency. The components of these indices, along with baseline means, are listed in Appendix A. The logic underlying this approach is that if the treatment affects health care delivery through its impact on intermediate outcome Q, then we should see an effect of the treatment on Q. Estimating treatment effects on these intermediate outcomes can thus help us gain a deeper understanding of the mechanisms through which the intervention operates.

[22] Further details of the procedures employed to ensure data quality are discussed in Appendix B.

[23] The baseline sample does not include the additional short surveys administered in the information and mobilization arms.

[24] As discussed in Appendix D, we supplemented this health center-level synthetic cohort data with a child-level measure that leverages the detailed child-month level retrospective data we collected at endline. Results for these child-level estimates are shown in Appendix E.

Table 1: Main outcome indices and their components

	Mean
Utilization	
Vaccination rates of children $<$ 36 months for polio, DPT, BCG, and measles, by age bracket $\star$	75.26%
Share of self-reported visits to the health center versus other providers	37.47%
Number of self-reported visits to the health center by household members in past 12 months	14.01
Treatment quality	
Whether equipment was used during the most recent visit	68.01%
Total time spent waiting for the initial consultation and the examination	104.28 mins.
Whether person seeking care was examined by trained health center staff during most recent visit	99.91%
Whether person seeking care had privacy during most recent examination	89.24%
Whether lab tests were administered during most recent visit	62.76%
Whether diagnosis was clearly explained to person seeking care during most recent visit	59.50%
Percent of staff in attendance during unannounced visit to health center	29.32%
Condition of health center (cleanliness of floors and walls; smell) as observed during unannounced visit	80.26%
Share of months in which stock cards indicate availability of six key tracer drugs in past 3 months, as determined during unannounced visit	93.15%
Patient satisfaction	
Whether services offered at health center are judged to be of "very"/"somewhat high" quality $\diamond$	45.89%
Whether person seeking care was "very satisfied"/"satisfied" with care received during most recent visit	67.77%
Whether person conducting examination appeared interested in health condition of person seeking care	90.08%
Whether person conducting examination listened to what person seeking care had to say	90.31%
Whether person seeking care felt free to express him/herself to person conducting examination	83.11%
Whether, compared to the year before, availability of medical staff has improved	48.76%
Health outcomes	
Weight for age among children aged 0-18 months	1.23
Weight for age among children aged 18-36 months	1.39
Upper arm circumference among children aged 0-18 months	2.51
Child mortality	
0 to 5 years (main measure)	0.05‰
0 to 12 months	0.04‰
1 to 5 years	0.01‰

$\star$ Vaccination rates are calculated at the household level as the percentage of children under 36 months who, subject to a six week grace period, have received the full set of age-relevant vaccinations as recommended in the Uganda National Expanded Program on Immunization.

$\diamond$ Baseline values for this variable were not collected; values shown are from the control group. The baseline index omits these components.

4.2.2 Missing Values and Outliers

As specified in our pre-analysis plan, we remove outliers by capping unbounded variables at the 99th percentile of the observed values in our data. To deal with missing values on our covariates, we adopt the approach described in Lin, Green and Coppock (2016). If no more than 10% of the covariate's values are missing, we recode the missing values to the overall mean. If more than 10% of the covariate's values are missing, we include a missingness dummy as an additional covariate and recode the missing values to zero. We deal with missing values on our outcome measures by setting them equal to the mean of the treatment group.

4.2.3 Social Desirability Bias and Hawthorne Effects

Although many of the outcomes we measure are products of objective observation by our survey team, several are based on subjective reports by household members. This raises the possibility that respondents might provide more positive answers to certain questions or report greater satisfaction with the quality of the services they received because they believe such answers will reflect better on them in the eyes of the interviewer.[25] While we cannot completely rule out such biases, we note that they should be balanced across treatment and control arms (since we collected outcome data in the same way in both), and hence should not affect our estimates of treatment impact. Furthermore, we took great care to decouple the intervention and the data collection exercise in the perception of respondents.

A greater concern is that health center staff and/or household respondents may have behaved or answered questions differently because they knew they were in the treatment group. In the case of health center staff, we believe we can largely rule out such Hawthorne effects because we implemented the health center survey—by far the most intrusive aspect of the intervention from the standpoint of the clinic staff—in both treatment and control units.

In the case of household members, concerns regarding Hawthorne effects are further minimized because only 20 percent of surveyed households in treated catchment areas reported having even heard about the CRC or the community or interface meetings. So, it is unlikely that knowledge of treatment status affected our estimates of program impact on outcomes related to citizens' behavior. We can also rule out the parallel concern that members of the survey team might have sought to validate the program's objectives through the way they asked questions or recorded observations about the clinics they visited because, as noted, survey team members were blinded to treatment status.

[25] It is also possible that the act of being surveyed may affect respondents subsequent health behaviors (Zwane et al., 2011).

4.3 Estimation

4.3.1 Main Effects and Intermediate Outcomes

To estimate the effect of the full treatment, we estimate the following ITT equation:

$$Y_{ij} = \beta_0 + \beta_1 T_{ij} + \beta_2 Y_{ij}^0 + \beta_3 X_{ij} + \beta_4 X_{ij} * T_{ij} + \phi_d + u_{ij} \quad (1)$$

where Y_{ij} is the outcome measure (in our main specifications, one of our five indices) of household i in health center catchment area j. T_{ij} is a binary variable indicating whether the health center and catchment area j was assigned to treatment. β_1 is the average treatment effect, Y_{ij}^0 is the baseline value of the outcome measure[26] X_{ij} is a vector of demeaned controls,[27] $X_{ij} * T_{ij}$ is their interaction with the treatment indicator,[28] ϕ_d are district fixed effects, and u_{ij} are robust standard errors clustered by the health center catchment area. For child mortality, the unit of observation is the health center catchment area. Following Lin (2013), we use Huber-White sandwiched standard errors.

We also use Equation 1 to estimate the effects of treatment on the intermediate outcomes described in Section 4.2.1.

4.3.2 Analysis by Treatment Arm

To test the effect of each treatment arm, we estimate the model:

$$\begin{aligned} Y_{ij} = \beta_0 &+ \beta_1 T_{ij}^{InfoOnly} + \beta_2 T_{ij}^{Info\&Interface} + \beta_3 T_{ij}^{InterfaceOnly} + \beta_4 Y_{ij}^0 + \beta_5 X_{ij} \\ &+ \beta_6 X_{ij} * T_{ij}^{InfoOnly} + \beta_7 X_{ij} * T_{ij}^{Info\&Interface} + \beta_8 X_{ij} * T_{ij}^{InterfaceOnly} + \phi_d + u_{ij} \quad (2) \end{aligned}$$

where $T_{ij}^{InfoOnly}$ is a binary variable indicating whether the health center and catchment area j was assigned to receive only the information treatment, $T_{ij}^{InterfaceOnly}$ indicates whether the unit was assigned to receive only the interface treatment, $T_{ij}^{Info\&Interface}$ indicates whether the unit was assigned to receive the full treatment, and all other terms are defined as in Equation 1. This set-up allows us to compare each cell in the factorial design to the control group.[29]

[26] We did not collect baseline values for a subset of index components, as highlighted in Tables 1 and A1. In these cases, the baseline value of the outcome index omits this component. For analyses of treatment effects on these individual components, the baseline value is omitted from the estimating equation.

[27] As specified in our pre-analysis plan, the controls include whether the health center is a HC2, whether the health center provides delivery services, whether the health center has staff houses, whether household members report using the health center within the 12 months prior to baseline, the education level of the interviewed household head, and household wealth (calculated as the first component of a principal component analysis of the number of items of 17 assets—including cattle, radios, bicycles etc.—owned by the household, as well as three measures of housing quality).

[28] The inclusion of the interaction between the controls and the treatment dummy was not pre-specified. We added this term in line with the recommendations in Lin, Green and Coppock (2016).

[29] We had initially pre-specified the model $Y_{ij} = \beta_0 + \beta_1 T_{ij}^{Info} + \beta_2 T_{ij}^{Info} T_{ij}^{Interface} + \beta_3 T_{ij}^{Interface} + \beta_4 Y_{ij}^0 + \beta_5 X_{ij} + \phi_d + u_{ij}$, which considers the rows and columns in Figure 2 as well as their interaction. We deem the model

4.3.3 Subgroup Treatment Effects

To test for subgroup treatment effects, we undertake a number of tests for treatment effects on both the five main outcome indices and the seven intermediate outcome indices in particular subsets of our sample. We estimate the standard equation:

$$Y_{ij} = \beta_0 + \beta_1 T_{ij}^k + \beta_2 T_{ij}^k * Sub_{ij} + \beta_3 Sub_{ij} + \beta_4 Y_{ij}^0 + \beta_5 X_{ij} + \beta_6 X_{ij} * T_{ij}^k + \phi_d + u_{ij} \quad (3)$$

where Sub_{ij} is an indicator variable of the subgroup for which we are testing for treatment effects, which for this purpose is not included in the vector of covariates X_{ij}.[30] β_2 is the marginal increase in the treatment effect in the health centers/catchment areas in this subgroup.

5 Results

As a first step, we check for covariate balance to ensure we are drawing inferences from valid comparisons. As shown in Appendix C, our sample is balanced across treatment arms with respect to the baseline characteristics of the catchment areas and health centers. Baseline levels of our main and intermediate outcome indices are also balanced. We test for evidence of treatment spillover by comparing outcomes in control health centers that were close to and far from the nearest treated health center, and find no statistically significant differences.[31] If anything, we find that the effect of exposure to the intervention on treatment quality is stronger in control health centers that are *further away* from treated units (see Appendix C).

5.1 Main Outcomes

Figure 3 presents the study's main findings. The coefficient plot summarizes the effect of the full ACT Health program on the five main outcome indices as measured at endline (20 months after the initial treatment). Corresponding regression tables for the outcome indices as well as their components (both standardized and non-standardized) are included in Appendix E.1. The dots represent the estimated treatment effect in standard deviation units; thin error bars represent the 95% confidence interval; thick error bars the 90% confidence interval. We find null effects on utilization rates, health outcomes, and child mortality but positive effects on the quality of care provided by health care providers and patient satisfaction, which increased by 0.070 and 0.077 standard deviations, respectively.

described in Equation 2 superior since it relies on fewer assumptions, is easier to interpret, and presents our findings in a way that is consistent with the results in the main specification. Results from the pre-specified model are presented in Appendix E.6.

[30] For specifications looking at subgroup effects by health center level we exclude the three health center level covariates from the vector since they have limited variation, leading to concerns about multicollinearity.

[31] "Close" control health centers are defined as those whose distance to the nearest treated health center was less than 5.2 miles, which is the 67th percentile of distances among all closest control/treated pairs in our sample.

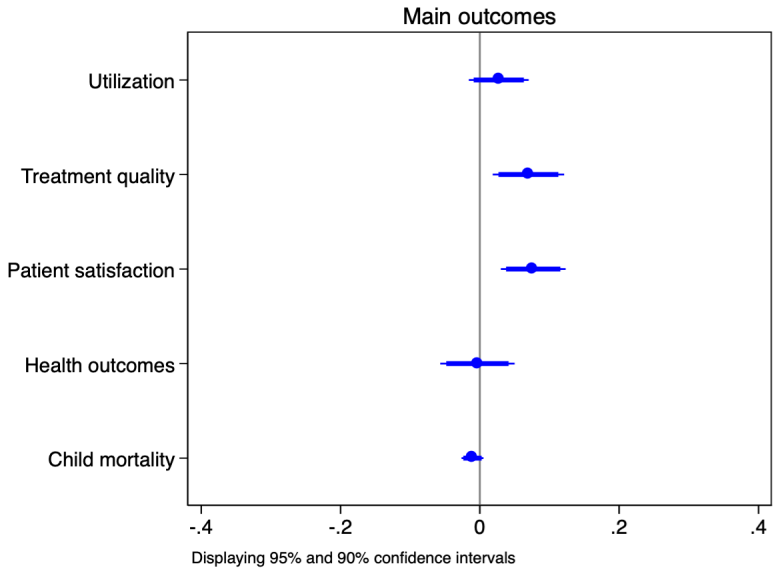


Figure 3: Effect of the full treatment at endline

We underscore that the substantive size of our estimates on treatment quality and patient satisfaction are not particularly large: our sample size puts us in a position to detect even small effects with confidence.[32] But they do speak to the positive effect of ACT Health on these two outcomes. Our null results on health outcomes and child mortality, on the other hand, are unambiguous, precisely estimated zero effects.

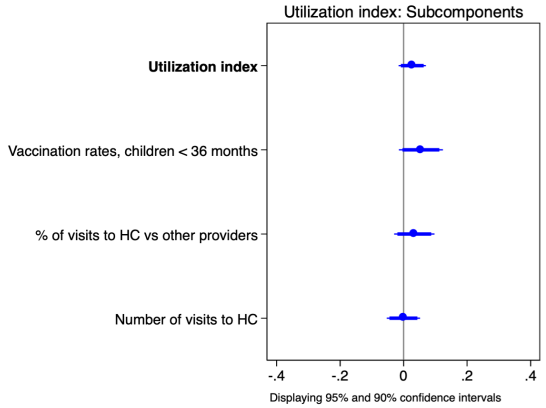
Figure 4 unpacks these index-level results into their components. As the Figure makes clear, the null findings with respect to utilization, health outcomes, and child mortality are rooted in statistically insignificant coefficient estimates on every index component. The one utilization index component that approaches conventional levels of statistical significance—and that, in fact, drives the nearly significant index coefficient—is child vaccination rates. The effect on the critical number of visits to the health center during the past 12 months measure, however, is a precisely estimated zero.

The patient satisfaction findings, by contrast, are a product of significant, positive estimates on every component but one (which is still positive, but not statistically significant). Households in treated communities were more likely to report at endline that the services offered at the health center were of "very" or "somewhat" high quality (5.2 percentage points); that they were "satisfied" or "very satisfied" with the quality of the care they received during their most recent visit to the clinic (2.3 percentage points); that the person conducting their examination behaved politely and respectfully (1.6 percentage points), appeared interested in their health condition (2.3 percentage points), listened to what they had to say (1.6 percentage points); that they felt free to express themselves to the person conducting the examination (1.1 percentage points, insignificant), and that, compared to the year before, the availability of medical staff had improved (3.5 percentage

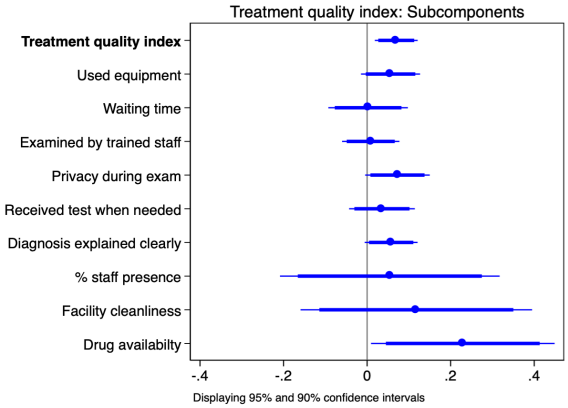
[32] We note as well that the sizes of these treatment effects are small relative to the secular changes taking place in several of our outcomes in *both* treatment and control units during the period we study.

Figure 4: Treatment effects on main indices and their subcomponents at endline

(a) Utilization

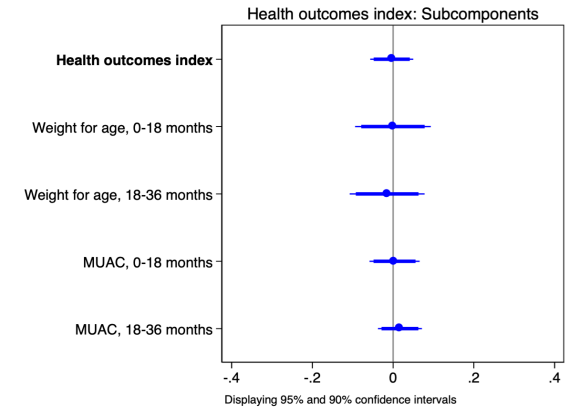


(b) Treatment quality

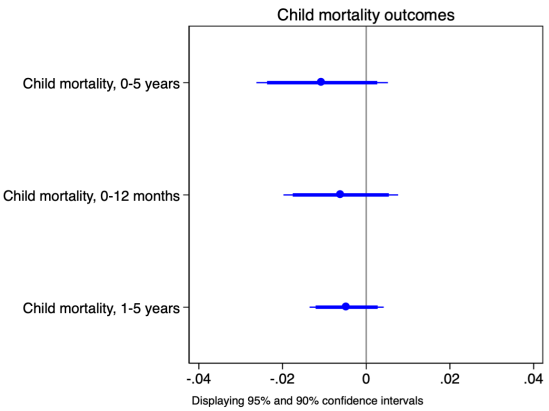


(c) Patient satisfaction

(d) Health outcomes



(e) Child mortality



points).[33]

[33]Since patient satisfaction can only be reported by households that utilized the health center during the past year,

Our significant results with respect to treatment quality are built on somewhat more mixed component-level findings. Respondents in households who received their care from treated health centers were more likely to report having had privacy during their most recent exam and having had their diagnosis clearly explained to them (1.5 and 2.4 percentage points, respectively). Treated health centers were also 6.2 percentage points less likely to have had stockouts of key drugs during the past three months. Although these three index components are the only ones for which treatment effects reach traditional levels of statistical significance, all of the other components also have positive coefficients, resulting in a significant positive estimate for the index as a whole. This positive index-level effect is robust to several alternative specifications, including (with one exception, discussed below) dropping index components one by one and excluding the three index components measured at the health center level (observed staff presence, cleanliness, and drug availability), whose inclusion in the household-level index artificially inflates their contributions (see Appendix E.4).

The only index component whose single omission causes the treatment quality index to lose its statistical significance is drug availability. Drug stockouts are more than just a statistically influential index component, however. The unavailability of essential medicines is a major source of poor health—and even death—in rural Uganda. Uganda employs a hybrid "push-pull" system under which requested quantities of basic drug supplies are sent to clinics from the National Medical Store (BMAU, 2015; Rwothungeyo, 2016). Hence, exposure to the ACT Health intervention might reduce stockouts via two channels: First, health workers who might otherwise file incomplete or late paperwork requesting drugs might be impelled by the complaints they hear from community members to project their drug needs more accurately and to request restocking in a more timely manner. Second, interacting with community members might cause health workers to resist the temptation to steal clinic drugs and sell them to patients at private pharmacies that they control or in which they have financial interests (Arinaitwe, 2017). Such drug thefts by clinic staff are a major problem in Uganda: 88 percent of households in our sample cited health workers selling drugs on the side as an important factor in explaining poor health service delivery. The problem is so severe that in 2009 President Museveni established a special agency within State House to combat the issue. The outsized contribution of drug availability to our treatment quality index can therefore be defended by pointing to the substantive importance of reducing drug stockouts to improving health outcomes.[34]

5.1.1 Midline Results for Main Outcomes

Our findings at midline are generally consistent with those at endline. As shown in Appendix E.7, we find no effects of exposure to the ACT Health intervention on utilization, health outcomes,

and since utilization is a post-treatment outcome, it is in theory possible that our estimates could be biased by a treatment effect on utilization. We think this is highly unlikely, however, given the statistically insignificant effects of the treatment on utilization. Furthermore, the treatment effect on patient satisfaction is robust to restricting the sample to the 92.5% of households who had visited the health center in the 12 months before baseline.

[34]The importance of drug stockouts as an outcome measure is also bolstered by the fact that our research staff measured drug availability directly through the physical inspection of each health center's pharmacy shelves during an unannounced visit. This makes our measure of drug stockouts immune to the subjective reporting that may possibly affect our other treatment quality index components.

or child mortality and a significant but substantively small (0.06 standard deviations) effect on treatment quality when we use outcome data measured 8 months, rather than 20 months, after treatment. In contrast to our endline findings, we observe no treatment impacts on patient satisfaction at midline. Exposure to the ACT Health intervention thus does not appear to have had shorter-term effects that dissipated by the time of our endline data collection.

5.1.2 Robustness Tests

In addition to the main results shown in Figure 3 and Appendix E.1, we find consistent effects in t-tests (see Appendix E.8), and in various alternative models we pre-specified in our pre-analysis plan. As we show in Appendix E.4, running the models without control variables or district fixed effects, aggregating all outcome measures to the health center level, and re-specifying our outcome measures as the difference between post-treatment and pre-treatment values all leave our findings substantively unchanged. We also show that our estimated null effects on child mortality are unchanged when we re-analyze our data using at the child level using a Cox proportional hazards model, leveraging the fact that we have child-month data on survival over the course of 36 months for over 20,000 children (again, see Appendix E.4).

To allay concerns that the number of hypotheses we test might lead us to falsely report statistically significant effects, we provide estimates of treatment impact on all indices and index components both with and without False Discovery Rate adjusted p-values (Benjamini and Hochberg, 1995), based on the comparison families described in Appendix E.9.

Quantile regressions of our five outcome indices (reported in Appendix E.4) suggest that our estimated treatment effects (both null and positive) are not driven by just parts of the distribution. Our results on utilization, patient satisfaction, and health outcomes are also robust to substituting our main pre-registered outcome measures with alternative indices based on the fist component of a principal component analysis (also see Appendix E.4). This is important insofar as our pre-registered indices, while deductively coherent, might not perfectly capture the underlying outcome they were designed to summarize.

5.2 Intermediate Outcomes

To better understand the channels through which the ACT Health intervention affected our outcomes of interest, we collected data on a range of intermediate outcomes. These include knowledge of patients' rights and responsibilities among community members; sense of efficacy among households; perceived community responsibility for monitoring health service delivery; monitoring activities undertaken by community members; and the perceived quality of community members' relationship with health care workers. In addition, we collected data on health workers' knowledge of patients' rights and responsibilities as well as actions the health center staff may have undertaken to improve transparency vis-à-vis the community (for example, having a suggestion box or posting opening times, a duty roster, and information about services provided and patients' rights). The components of each of these indices are listed in detail in Appendix A.

As can be seen in Figure 5, we do not find evidence for positive treatment effects at endline on any of the intermediate outcome measures we collected (see Appendix E.2 for regression tables).[35] The fact that we see no impact on efficacy, community responsibility, or community monitoring—including in indices constructed using principal component analysis rather than averaged z-scores (see Appendix E.4)—is especially noteworthy, as these are the three indices that speak most directly to the role that citizens may play in generating bottom-up accountability. Given the expectation that information provision will affect health provision through its impact on citizen monitoring and bottom-up pressure, this is an important result. It suggests that the modest improvements we observe in treatment quality—and ultimately patient satisfaction—may have been driven less by the effect of the intervention on community actions than by a direct effect on health workers' behavior.

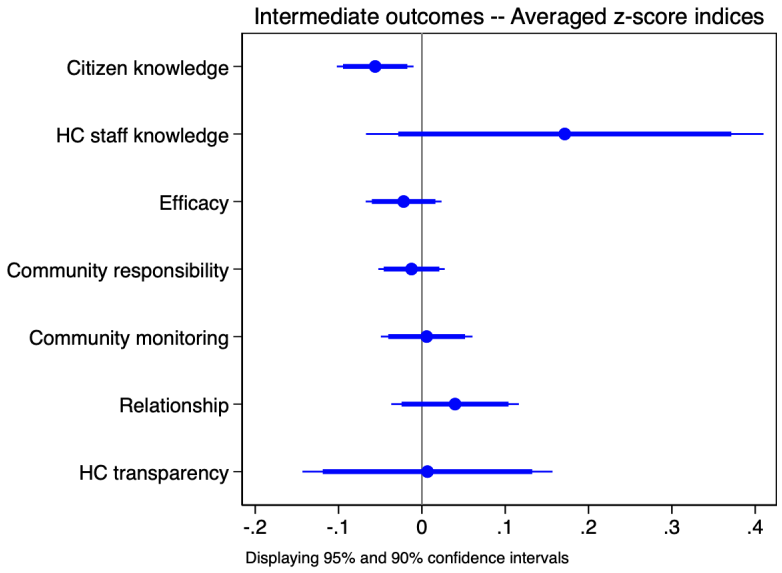


Figure 5: Treatment effect on intermediate outcomes at endline

5.3 Differences by Treatment Arm

Our motivation for the factorial research design was to disentangle which aspects of the bundled full intervention were doing the work, if any. Table 2 shows effects on the five main outcomes by treatment arm. Each of the three treatment arms enters as an indicator variable.[36]

[35] Insofar as citizen knowledge can be thought of as a manipulation check in an information-focused intervention like ACT Health, the significant negative sign on that intermediate outcome measure may appear troubling. We note however that the estimate loses significance once a multiple testing adjustment is applied and that the substantive size of the coefficient is, in any case, tiny—corresponding with an additional fraction of a right or responsibility correctly named by respondents in the control group.

[36] We present the treatment arm-level results for each index component and for our intermediate outcomes, alongside results from our pre-specified model, in Appendix E.6. The midline results for the treatment arm-level analysis are presented in Appendix E.7.

The null effects on utilization, health outcomes, and child mortality in the full treatment are generally unchanged in the information/mobilization and interface sub-treatments, although we do see a small, positive effect on utilization rates (0.05 standard deviations, significant at the 95% level) in the interface only arm and a small reduction in child mortality rates (0.02 standard deviations, significant at the 95% level) in the information and mobilization only arm.[37] Both sub-treatments have similar positive effects on patient satisfaction. The effect on treatment quality is marginally positive in all arms, but only significant in the full treatment arm.

Table 2: Main outcomes – All treatments

	(1) **Utilization**	(2) **Treatment quality**	(3) **Patient satisfaction**	(4) **Health outcomes**	(5) **Child mortality**
Full treatment	0.027	0.071***	0.080***	-0.003	-0.011
	(0.022)	(0.026)	(0.024)	(0.028)	(0.008)
Information and mobilization only	0.013	0.013	0.073***	-0.023	-0.020**
	(0.022)	(0.029)	(0.026)	(0.029)	(0.008)
Interface only	0.054**	0.022	0.064***	-0.011	-0.009
	(0.022)	(0.027)	(0.022)	(0.028)	(0.008)
Constant	-0.018	-0.002	-0.006	-0.488***	0.061***
	(0.015)	(0.021)	(0.018)	(0.022)	(0.006)
N	14,609	14,609	14,609	10,023	376
R^2	0.221	0.102	0.040	0.103	0.151
P-value (Info/mobilization = Interface)	0.066	0.740	0.697	0.653	0.204
P-value (Info/mobilization = Full treatment)	0.518	0.025	0.778	0.472	0.281
P-value (Interface = Full treatment)	0.234	0.032	0.441	0.767	0.820

Notes. Estimates comparing outcomes between each treatment arm and the control. Each treatment arm enters as an indicator variable. Models (1)-(5) include district fixed effects as well as demeaned baseline covariates and their interactions with the treatment indicators. *** p<0.01; ** p<0.05; * p<0.10

5.4 Subgroup Effects

The evidence presented thus far speaks to the impact of ACT Health *on average*. However, it is possible that the intervention may have had different effects in subsets of health centers and catchment areas with different characteristics. Investigating such sub-group effects can be helpful for better understanding the mechanisms at work and for generating expectations about the likely

[37]This latter effect is consistent with the findings in Björkman Nyqvist, de Walque and Svensson (2017), which suggest that information is a necessary ingredient for bottom-up accountability to work. However, given both the null effects of the information/mobilization treatment on health outcomes and utilization and the statistically insignificant differences across the information/mobilization, interface, and full treatment arms (see bottom panel), we hesitate to read too much into this finding.

external validity of the findings in other settings and populations (Banerjee, Chassang and Snowberg, 2017). Table 3 summarizes the results of our investigation into such differential effects. A "0" indicates that the intervention did not have a significant effect in the specific subset of health centers or catchment areas specified at left. A "+" indicates that the intervention had a positive treatment effect, a "-" indicates that it had a negative treatment effect, both at the 90% significance level or greater. Levels of significance are indicated by the shade of gray, with dark gray indicating 99% significance, medium gray 95% significance, and light gray 90% significance. The specific coefficient estimates on which these codings are based are shown in Appendix E.3.[38]

The results of our investigation into subgroup effects reinforce our null findings with respect to utilization, health outcomes and child mortality by demonstrating that these null results hold across nearly all subsets of health centers, catchment areas and households. Our positive findings with respect to treatment quality and patient satisfaction, meanwhile, are bolstered by the fact that, as we describe below, we are more likely to find effects in places where theory suggests the intervention should have had the greatest impact.

5.4.1 Health Facility Characteristics

The first health facility characteristic we test is the health center's level. As noted in Section 2, HC3s and HC2s have different kinds of personnel and offer different services. In addition, the types and sizes of the communities that HC3s and HC2s serve also differ considerably: the former typically provide health care for a subcounty with a median population of 20,000 people, whereas the latter typically serve one or two parishes, or about one- to two-fifths of a subcounty. All of these factors could lead to divergent treatment effects. While we find no effects with respect to utilization, health outcomes or child mortality across either HC2s or HC3s, we find positive and significant effects on patient satisfaction in both types of health centers. Effects on treatment quality, meanwhile, are significant only in HC2s.

The second clinic-level characteristic we address is whether the health center is performing above or below the median level in its district on our treatment quality index, as measured at baseline. The rationale for this test is that the impact of the intervention may be different in well-performing and poorly-performing health centers—in part because the nature of the information contained in the CRC (and thus the treatment itself) will be different and in part because different baseline conditions imply varying degrees of room for improvement. Consistent with this expectation, we find larger increases in treatment quality and patient satisfaction in health centers with low baseline performance.[39]

[38] Appendix E.3 presents results of the subgroup analysis for our intermediate outcomes.

[39] We also find increases in treatment quality and patient satisfaction in health centers with high baseline performance, but these increases are smaller, and are correspondingly significant at a lower level.

Table 3: Subgroup treatment effects on main outcomes

		(1) **Utilization**	(2) **Treatment quality**	(3) **Patient satisfaction**	(4) **Health outcomes**	(5) **Child mortality**
Health center level	HC2	0	+	+	0	0
	HC3	0	0	+	0	0
Treatment quality	High	0	+	+	0	0
	Low	0	+	+	0	0
Alternative health care options	High	0	0	+	0	0
	Low	+	+	0	0	0
Embeddedness of HC staff	High	0	+	+	0	0
	Low	0	+	0	0	0
Collective action potential	High	0	0	+	0	0
	Low	0	+	+	0	0
Community monitoring	High	0	0	+	0	0
	Low	0	+	+	0	0
Efficacy	High	+	+	+	0	-
	Low	0	+	+	0	0
Health NGOs present in the village	Yes	0	0	+	0	0
	No	0	+	0	0	0
Avg. distance to HC	High	0	+	0	0	0
	Low	+	+	+	0	0
Catchment area	Urban	+	0	0	0	0
	Rural	0	+	+	0	0

Notes. This table shows estimated average treatment effects for subgroups of health centers. Each pair of subgroup effects is derived from a separate regression, estimated using Equation 3. The table displays the coefficient on Full Treatment for the base category (such as for example HC2) and the linear combination of the coefficient on Treatment and on the interaction between Full Treatment and an indicator variable describing the specific subset (such as for example Treat + Treat * HC3 for HC3). For continuous variables, *High* indicates that a health center's value for the given variable is at or above the median, *Low* indicates that it is below the median. + indicates a positive coefficient significant at least at the .1 level, - indicates a negative coefficient significant at least at the .1 level, and 0 indicates an insignificant coefficient. A dark shade of gray indicates significance at the 99% level, medium gray at the 95% level, and light gray at the 90% level.

A third facility-level characteristic we investigate is the availability of alternative health care options. The more numerous the alternative sources of health care, the greater the likelihood that community members will respond to the receipt of information about poor service provision at their own health center by exiting rather than exercising voice. This leads us to expect stronger

treatment effects in health centers with fewer nearby alternative health care options (which we measure in terms of the share of self-reported visits to the sampled health center versus other government health centers, private health centers, or traditional healers at baseline). In keeping with expectations, we find that exposure to ACT Health is associated with improvements in treatment quality, and also increases in utilization, only in places where people have fewer alternative health care options. Patient satisfaction, meanwhile, is affected by treatment only in areas where alternative health care options are more plentiful (perhaps because patient satisfaction is only recorded among households that use the health center, and those that are dissatisfied select out).

Finally, we test whether the extent to which health workers are embedded in the communities they serve—attending church there, sending their children to school, living there—moderates treatment effects. In line with Tsai (2007), we may expect health workers who are more embedded in local social networks to be more susceptible to informal pressure and thus more likely to improve their behavior in response to the community dialogues and interface meetings. We find mixed support for this hypothesis: we find statistically significant increases in patient satisfaction only in health centers where health workers are more embedded in the community, while treatment quality increased regardless of the degree to which staff are embedded.

5.4.2 Catchment Area Characteristics

In addition to these facility-level attributes, we also test for varying treatment effects across catchment areas with different characteristics. The first of these is the catchment area's collective action potential at baseline. This is likely to be important insofar as the intervention depends on the ability of community members to work together to monitor health center staff and sanction them if they are found to be underperforming. We measure the community's collective action potential by constructing a z-score index of two components: ethnic homogeneity and the share of people who say they believe they could mobilize members of their community to press for improved health care. The rationale for the first component stems from the large literature indicating that ethnically diverse communities have a more difficult time achieving collective ends (e.g., Miguel and Gugerty (2005), Khwaja (2009), Algan et al. (2016)). Consistent with the findings in this literature, Björkman and Svensson (2010) report that the impact of the P2P intervention was stronger in health centers that were located in more ethnically homogeneous districts. We, however, find that exposure to the intervention has significant effects on treatment quality only in catchment areas with *lower* collective action potential.[40] Patient satisfaction, meanwhile, increases in areas with both high and low collective action potential.

Insofar as ACT Health is about community monitoring, we might expect communities that are already actively engaged in monitoring to respond more strongly to the intervention than communities whose baseline levels of monitoring are lower. On the other hand, to the extent that ACT Health does in fact generate community monitoring, we might expect treatment effects to be larger in communities where such monitoring is not already taking place. We test these expectations by

[40]One reason for the different findings in Björkman and Svensson (2010) and in our own analysis is that we are measuring ethnic heterogeneity at different levels. Björkman and Svensson (2010) measure diversity at the district level, whereas we measure it at the level of the catchment area—the more relevant unit if we want to capture the collective action potential of the community whose mobilization may affect the health center staff's behavior.

comparing outcomes in communities whose baseline community monitoring index values (as defined in Appendix A) are above and below the median. In keeping with the second hypothesis, we find that treatment quality increases only in settings where baseline community monitoring activity is low. Patient satisfaction, meanwhile. increase in both types of communities.

Another catchment area characteristic that may affect treatment uptake is the community's baseline level of efficacy. As Lieberman, Posner and Tsai (2014) argue, even individuals who are motivated to act by the receipt of information about poor service delivery may be dissuaded from mobilizing if they do not believe they have the power to effect change. The implication is that we would expect communities in which baseline levels of efficacy (as measured by our efficacy index, described in Appendix A) are above the median to respond more strongly to the intervention. On the other hand, to the extent that programming in ACT Health is meant to be efficacy-boosting, we might expect to see communities with lower baseline levels of efficacy exhibiting the strongest treatment effects. We find that both treatment quality and patient satisfaction increased irrespective of the baseline level of efficacy in the community. In places with high baseline levels of efficacy, we also find significant positive treatment effects on utilization rates.

The intensity of health-oriented NGO activity in the catchment area might also condition the impact of ACT Health, although it is not theoretically clear whether the presence of other NGOs should attenuate our estimates of treatment effects (via diminishing returns of exposure to NGO health programming) or amplify them (by reinforcing the impact of exposure to the new intervention). We address this issue by leveraging answers to a question asked in a survey of local council (LC1) chairs about whether there were other NGOs in the village dealing with health issues. Consistent with the hypothesized attenuating effect of exposure to other NGO programming, we find that the positive impact of ACT Health on treatment quality is only significant in areas in which other health-oriented NGOs are absent. Patient satisfaction effects, meanwhile, are found only when NGOs are present.

Finally, it is reasonable to expect the intervention to be more likely to have an impact in areas where the average distance between households and the health center is relatively low, since this will make it easier to gather and share information and to check on the facility and its staff. For similar reasons, we might also expect to find stronger effects in areas that are more urban. Consistent with these expectations, we find evidence for treatment effects on utilization in urban areas and in places with a low distance between households and the health center. Effects on treatment quality and patient satisfaction, meanwhile, are driven by the intervention's impact in more rural areas and places with a low distance between households and the health center.

6 Discussion

The primary objective of our evaluation was to test whether the ACT Health program—and, by extension, interventions like it that aim to improve local service delivery through information provision and community mobilization—generated improvements in the health outcomes of citizens living in proximity to treated clinics. Although we do find some evidence for the program's effects on the quality of care provided at those clinics, we find no impact on health outcomes *per se*. These

results are robust to numerous alternative specifications, including an exploration of treatment effects in subsets of health centers, communities, and households where theory leads us to expect to find stronger effects of exposure to the intervention.

Beyond these main results, two other findings have important implications for the literature on service provision and accountability. The first is our finding that exposure to the intervention causes patients to say they are more satisfied with the quality of the care they receive at their local health center. The second is the lack of evidence we find that the intervention had any effect on citizens' monitoring behavior. We discuss each in turn.

6.1 Patient Satisfaction

In light of the evidence that ACT Health led to improvements in treatment quality—and that these improvements were already apparent by midline (see Appendix E.7)—our findings with respect to patient satisfaction make sense: the increase in patients' satisfaction with their care is plausibly a response to the positive changes in health providers' behavior, as captured in our treatment quality index.[41] Since these changes in provider behavior were not associated with measurable changes in actual health outcomes, we can infer that patient satisfaction may be rooted in the character of patients' interactions with their health care providers rather than in improvements in health outcomes that these interactions may generate.[42]

An alternative interpretation is that our findings on patient satisfaction are due less to changes in health provider behavior (which, after all, are substantively quite small) than to the participatory nature of the ACT Health intervention. Other studies have found similar increases in citizen satisfaction following community members' participation in interventions that involve consultation and/or direct participation in decision-making, even when the interventions have no tangible effects on other outcomes. For example, Olken (2010) finds that Indonesian villagers whose communities were randomly assigned to choose local development projects by direct plebiscite rather than through meetings of representative councils were much more satisfied with the outcome of the process, even though the projects that were ultimately selected were no different. Beath, Christia and Enikolopov (2017) find similarly that citizens in Afghanistan who participated directly in the selection of local development projects were significantly more satisfied than those whose projects were chosen by elected village elites, even when the projects selected were equally in keeping with their preferences.

These findings suggest that including non-elite community members in decision-making processes can generate satisfaction with the outcomes generated, even if the outcomes themselves are unaffected by the community members' participation. These effects may be particularly strong in settings like Indonesia and Afghanistan—and also Uganda—where, for reasons of elite capture and status differentials between regular citizens and service providers, community members rarely have their opinions taken seriously by elites and are ordinarily shut out of participation in collec-

[41] Consistent with this interpretation, we find suggestive evidence that changes in treatment quality between baseline and midline are associated with changes in patient satisfaction between midline and endline (see Appendix E.5). However, as seen in Table 3, this tracking does not hold within all subgroups.

[42] This is a common finding in the medical literature. For example, see Kahn et al. (2015).

tive decision-making. In such contexts, simply being asked for one's views and being in a position to interact on an equal basis with comparatively high status service providers (even if just in a facilitated meeting) may alter citizens' subjective perceptions of the performance of the actors and institutions that they are later asked to evaluate.[43]

An interesting wrinkle in our patient satisfaction findings is that they are absent at midline (see Appendix E.7). In contrast to the findings in Beath, Christia and Enikolopov (2017), who report that satisfaction decays during the two years between their midline and endline surveys, we find that patient satisfaction takes time to grow. This fact holds out the possibility that there might be longer-term consequences for health outcomes that we are not (yet) in a position to measure: if patients who are satisfied with the quality of the care they receive from their health center are more likely to seek treatment when they are ill—recall that, at baseline, 33 percent of households reported that they self-treat)—then improvements in patient satisfaction may lead to increases in downstream utilization that, in turn, may lead to improvements in health outcomes in the future. This possibility is consistent with the fact that while our estimate for the impact of exposure to ACT Health on utilization rates is still below the threshold of statistical significance at endline, it is greater than at midline (point estimate=0.027 at endline versus -0.012 at midline).

6.2 No Evidence of Community Monitoring

Equally important as our null findings with respect to utilization, health outcomes, and child mortality is the lack of evidence we find for treatment effects on community monitoring. In the literature on transparency and accountability, the whole rationale for providing information to citizens is that it will put them in a better position to monitor and sanction underperforming service providers (Khemani, 2007; Mansuri and Rao, 2013; World Bank, 2016). The ACT Health intervention, which included not just information provision but also hands-on mobilization of community members to encourage them to use the information they received to scrutinize the behavior of their local health providers and hold them accountable for poor performance, provides a strong test of this presumed causal channel. Our finding that ACT Health had no impact on any of our measures of community monitoring, but that treatment quality nonetheless modestly improved in health centers exposed to the intervention, therefore raises questions about the salience of this broadly accepted mechanism.

The lack of evidence for community monitoring comes not just from the null effects on the three intermediate outcome indices that capture aspects of community monitoring (as highlighted in Section 5.2) but also from the specific index components that most directly measure citizens' abilities to monitor and apply bottom-up pressure on health workers (see Appendix E.2 for details). For example, the household questionnaire asked respondents whether they thought that engaged

[43]This logic raises the possibility that the treatment quality effects we estimate might be driven by respondents' rosier views of health providers stemming from their pleasure at having been included in the study. We believe we can rule this out for two reasons. First, four of the nine components of the treatment quality index (including the crucial drug availability measure) were measured directly by our research team during its initial unannounced visit to the health center, and are thus not susceptible to reporting bias by respondents. Second, as discussed in Section 4.2.3 above, 80 percent of our household survey respondents reported that they had not participated in the community or interface meetings, and could therefore not have been affected—at any rate not directly—by having been included in the deliberative process or being treated as equals with higher status doctors and nurses.

community members would find out if a health worker did not report for work or did not provide the effort that he/she should in caring for patients. Seventy-three percent of households in the control group answered these questions in the affirmative, suggesting that citizens' confidence in their ability to detect poor health service delivery was already fairly high prior to the intervention.[44] However, among those exposed to the intervention, we see no increase in this confidence. Similarly, while 45 percent of household respondents at baseline reported that they thought they were responsible for making sure health workers came to work and provided high quality health services, exposure to the ACT Health programming generated no increase in this sense of responsibility for monitoring. Survey respondents in households located in treated villages were slightly more likely to report at midline that they thought they had a say in how health centers provided health care to their community, that they could pressure a health worker to report to work on time if the worker were regularly coming late, and that they could pressure a health worker to exert better effort in caring for patients. However, these effects disappeared by endline. To the extent that sustained confidence in one's ability to effect change is a necessary condition for citizens to invest in applying bottom-up pressure on health providers, these findings may help account for why we see such weak effects on citizen monitoring—and also why health center staff in treated and control units reported no differences in the rates at which community members called for meetings with health workers, made suggestions, or lodged formal complaints.

Notwithstanding the theoretical and policy appeal of the community monitoring approach, bottom-up pressure is extremely difficult to mobilize. Collective action problems may simply be too hard to overcome; citizens' efficacy and sense of responsibility for monitoring health care providers may be too weak; formal institutions such as local councils may be moribund and/or corrupt, and therefore unable to support citizens' monitoring efforts; and, compared to the other more immediate problems people face, health care may be insufficiently important to justify the investments in time and energy that the monitoring approach assumes community members will be willing to make to try to effect change (Lieberman, Posner and Tsai, 2014).

Although commonly invoked to provide theoretical justification for a bottom-up, information-focused approach to improving service delivery, the logic of the principal agent framework also helps explain the limits of such a strategy. As explicated in the classic theoretical treatments of Ross (1973), Arrow (1974), and Holmström (1979), and more recently summarized in Besley (2007), the crux of the principal-agent problem lies in two inherent characteristics of the relationship between any actor and the agent to whom she has delegated responsibility for completing a task. The first is that the principal cannot directly observe the actions of the agent—whether he comes to work on time (or at all), how hard he works, whether he has been wasteful with resources, etc. The second is that the outcome the principal observes is affected by factors outside of the agent's control. This makes it very difficult for the principal to make a clear inference about the agent's actions from the outcome that she observes (whether the task is completed expeditiously and with what quality). Simply supplying community members with information about the outcomes that have been achieved at the health center or how these outcomes compare with district averages (which is precisely the kind of information the CRC provides) does nothing to solve the problem of the health workers' effort being unobservable. If outcomes are found to be deficient, it will be difficult for community members to discern whether the poor performance stems from low

[44]These questions were not asked at baseline. The 73 percent figure comes from the control group at midline.

effort by the health center staff or, as the health workers will certainly claim, from circumstances outside of their control—underfunding, staff shortages, delays in the delivery of drugs and other supplies, or other factors. The provision of information may aid community members and health center staff in developing joint action plans that are built around problems over which they actually have control, as Björkman Nyqvist, de Walque and Svensson (2017) emphasize. But information alone will be insufficient for enforcing the agreements that those action plans contain.

These considerations are reinforced by the absence in the setting we study of another key factor stressed in principal-agent models: the ability to sanction. To the extent that information provision works, it may be that it only does where citizens have actual leverage over the frontline service providers they are being encouraged to monitor. In our study context, as in many developing country settings where interventions like ACT Health have been deployed, it is difficult to imagine how even highly mobilized citizens would be able to sanction underperforming service providers.[45] Absent the ability to sanction, investments in monitoring may appear futile, and thus not be made. Of course, service providers may alter their behavior in *anticipation* of citizen pressure, even if such pressure never materializes.[46] But such a response is not likely to be sustainable once it is revealed that sanctions are not forthcoming.

If citizens lack the power to sanction frontline health care providers from the bottom up, what about the local government officials who oversee them? Can this alternative set of principals, who by virtue of their formal oversight role and their connections with actors higher up in the government *do* have the ability to sanction underperforming health care workers, successfully apply pressure from the top down? And might such top-down monitoring bolster the efficacy of the bottom-up pressure that citizens have difficulty generating on their own? Although ACT Health did not explicitly involve district- or subcounty-level government health officials in its programming, such officials were informed of the intervention and invited to attend the community and interface meetings, and our implementing partners kept careful records of whether or not such officials did, in fact, attend these meetings (see Appendix H.2). Where they did, the effect of the intervention on treatment quality *nearly doubled* (see Appendix E.5). While the fact that the subcounty officials' attendance at the meetings was not randomly assigned cautions against reading too much into this finding, the result is highly suggestive of the power of top-down monitoring—perhaps in combination with bottom-up monitoring by citizens—to improve the performance of frontline service providers.[47]

While our results suggest that the ACT Health intervention did (at least modestly) change the

[45]Citizen monitoring interventions aimed at shaping the behavior of elected officials, over whom citizens in principle have sanctioning power via their votes, may be more promising (World Bank, 2016). Consistent with this argument, Grossman and Michelitch (2018) find that Ugandan politicians about whom performance information was circulated to voters did in fact perform better, but only in competitive constituencies where citizens possessed real leverage over the politicians. For a less optimistic set of findings about voters' sanctioning power, see Dunning et al. (Forthcoming).

[46]Indeed, Grossman and Michelitch (2018) identify precisely this type of anticipatory response, not the actual withdrawal of support by disenchanted voters, as responsible for the effect of information provision on politicians' behavior in their study.

[47]Raffler (2017) finds similarly that Ugandan councilors are effective in monitoring local bureaucrats only when they can call on Members of Parliament or local media as external levers. The findings are also consistent with World Bank (2016), which concludes that "a key condition for successful local action is the signal by higher-tier government departments that local citizen action would be taken seriously and used by leaders with power...to hold local officials accountable" (p. 199).

behavior of health care providers, they also suggest that these changes were not due to bottom-up pressure from citizens. It may be satisfying from a democratic theory perspective to think that we can "harness transparency and citizen engagement" (to borrow the sub-title of World Bank (2016)) or that the answer to the problem of poor service delivery lies in giving "power to the people." But mobilizing citizens to monitor and apply pressure on frontline providers may not be the most powerful strategy for improving the quality of health care and other services. Our findings suggest that more direct engagement with service providers, and perhaps top-down monitoring by government officials, may be a more promising approach.

7 Reconciling the Divergent Results in ACT Health and P2P

How can we reconcile the divergent findings in ACT Health and the original P2P study? Whereas Björkman and Svensson (2009) report strong effects on citizen monitoring, treatment practices, immunization rates, utilization, and health outcomes, we find (quite modest) effects only on treatment quality, and no statistically significant impact on utilization, health outcomes or child mortality.[48] We discuss several possible sources of these different findings.[49]

7.1 Different Samples

A first possible answer lies in the fact that P2P and ACT Health were implemented in different parts of Uganda, with districts from the Central Region overrepresented in the P2P sample (4 of 9) and underrepresented in the ACT Health sample (just 2 of 16—see the map in Appendix G).[50] Although both studies block randomize their treatment assignments by district to ensure that district-level variation in local conditions will not bias estimates of treatment effects, the different districts included in each sample might account for the differences in the estimated treatment impacts across the two studies. We investigated this possibility by sub-setting our analyses by region and find that our null results on utilization, health outcomes, and child mortality are found in all four regions (results not shown). This suggests that the contrasting findings in P2P and ACT Health are not likely to stem from the different districts in which the two programs were implemented.

Another difference between P2P and ACT Health is the levels of health centers included in each study. Whereas ACT Health was implemented in both HC3s and HC2s, P2P was implemented only in HC3s. As discussed in Section 5.4.1, health centers at these two levels differ in several ways that may affect treatment uptake. To test whether the inclusion of HC2s in our sample might be responsible for the two studies' divergent findings, we re-ran our analysis in HC3s only and find that our null results on utilization, health outcomes, and child mortality are unchanged (see Appendix E.10).

[48] We also find significant, albeit substantively modest, effects on patient satisfaction, but this is not an outcome that Björkman and Svensson (2009) measure.

[49] The differences highlighted in the discussion that follows are summarized in Appendix G.

[50] Of the 25 districts included in the two studies—9 in P2P; 16 in ACT Health—there was only one overlapping district, Apac.

7.2 Program Implementation and Data Collection

A second set of possible explanations lies in subtle differences in the ways the two studies were implemented. One possibility that we believe we can rule out with confidence is that the largely null results in ACT Health might be a product of poor program implementation. GOAL Uganda went to great lengths to train and monitor its local partners, and collected copious data on the implementation of its programming.[51] These data confirm that the programming was carried out faithfully.

Another possible explanation lies in the different ways that P2P and ACT Health worked with implementing partners to run the two interventions.[52] P2P was implemented by 18 community-based organizations (CBOs) that had been active, primarily in health-related programming, in 64 percent of the treatment communities and 48 percent of the control communities prior to the intervention (Björkman and Svensson, 2009). The CBOs' prior work in the P2P treatment areas may have bolstered the intervention's impact for several reasons. First, the social connections that CBO members had fostered with the community during their prior work may have made their mobilization efforts under P2P more effective and potentially increased community members' confidence that they could sanction underperforming health providers. Second, the CBOs' prior work in the area may have increased the communities' collective action potential, similar to the findings in Fearon, Humphreys and Weinstein (2015). Third, the specifically health-related programming undertaken prior to the P2P intervention may have laid the groundwork for the programming in P2P by familiarizing community members with the services offered at the local clinic, the importance of preventative health care practices, or other health promoting information. Finally, the CBOs may have chosen to work in the first place in communities where citizens and health providers were thought to be more responsive. Donato and Garcia Mosqueira (2019) explore this possibility in their replication and extension of Björkman and Svensson's analysis.[53] They show that the addition of a control for prior CBO presence and an interaction between prior CBO presence and treatment assignment leads to a reduction in the magnitude of some of the estimated effects reported in P2P. Their findings lend credence to the possibility that the P2P results may have been affected by the imbalance across treated and control units in the places where CBOs had previously worked.[54] ACT Health, by contrast, was implemented in partnership with four organizations that had not previously worked in the villages in which they were deployed.[55]

Another potential explanation for the differences in findings lies in the different amount of time the implementing teams spent in each community during their first mobilization visits (one half-day in ACT Health versus two half-days in P2P), and in the average number of participants

[51] For further details of GOAL's monitoring efforts, see Appendix H.2.

[52] For a more general discussion of the impact that different implementation partners may have on outcomes, see Banerjee, Chassang and Snowberg (2017).

[53] P2P was one of 35 influential studies for which the International Initiative for Impact Evaluation (3ie) sponsored internal replications (Brown and Wood, 2019). Internal replications use data from the original study in order to check the validity and robustness of its estimations and recommendations.

[54] For a simulation-based explanation for why including the CBO presence indicator and its interaction with the treatment dummy improves inferences in most scenarios, see Blair et al. (Forthcoming).

[55] We note that locating interventions in areas where implementing partners have prior experience is common practice in development programming. However, this design choice has implications for external validity and for scalability to settings beyond those where CBOs had previously worked (Alcott, 2015).

in the community meetings (100 in ACT Health versus 150 in P2P). Such reductions in the intensity of the intervention are common when programs are scaled up (Bold et al., 2013; Grossman, Humphreys and Sacramone-Lutz, 2016; Banerjee et al., 2017), and it is possible that this reduction in meeting size and time spent in the community may have generated weaker treatment effects on health outcomes in ACT Health. However, while P2P may have been more intense up front, ACT Health's longer timeframe allowed for three follow-up meetings prior to endline data collection (with average meeting size of 41 participants) compared to just one follow-up meeting in P2P. So there are also reasons for expecting ACT Health to have had a greater impact than P2P.

Finally, the P2P intervention, but not ACT Health, included a role playing exercise as part of the interface meeting in which community members and health center staff pretended they were in each others' positions. To the extent that such an exercise creates opportunities for discussing sensitive issues such as absenteeism or fosters a deeper appreciation of the challenges faced by the other actor in the principal-agent relationship, its absence from the ACT Health programming may have reduced its impact.

7.3 Variable Operationalization and Econometric Specifications

A third possible source of the divergent findings in ACT Health and P2P lies in the different variable operationalizations and statistical tests used in the two studies. We adopted somewhat different outcome measures and made slightly different modeling choices where we felt that modifications to the empirical strategy employed in P2P would generate stronger tests, or in order to better align our analyses with best practices in the public health field.[56] The inclusion of HC2s in our study also required changes to some of our outcome measures.[57]

To test whether our contrasting findings stem from these differences, we used Björkman and Svensson's own specifications and variable operationalizations to replicate the main tables in Björkman and Svensson (2009) using our own midline and endline data.[58] As we show in Appendix F, we are unable to replicate the P2P findings with respect to monitoring and information, immunization, utilization/coverage, or health outcomes. We infer that the differences between the results on these outcomes in ACT Health and P2P are not due to the different specifications or variable definitions used in each study.

[56] An example of the latter is our decision to measure child mortality using a life table, rather than vital statistics, approach (see Appendix D for details).

[57] Since delivery services are generally not provided at HC2s, we dropped delivery and antenatal visits from our utilization index to create a measure that could be applied across all units. We show in Appendix E.5 that adding these components back into our utilization index does not change our null findings with respect to utilization.

[58] Our analysis is based on the code provided to us by Katherine Donato and Adrian Garcia Mosqueira (the replication team that 3ie sponsored to replicate P2P), which they received from Björkman and Svensson. We were not provided access to Björkman and Svensson's original code.

7.4 Different Baselines

Perhaps the most powerful explanation for the different results in P2P and ACT Health lies in the ten year interval between the two interventions—the former in 2004-2005; the latter in 2014-2016. Health outcomes in Uganda improved markedly during this period. Under-five mortality fell from 117 to 59 per 1,000 live births; the share of births attended by a skilled provider rose from 42 percent to 74 percent; the incidence of stunting fell from 38 percent to 29 percent; wasting fell from 6.1 percent to 3.6 percent; and the share of children aged 12-23 months receiving all of their basic vaccinations rose from 46 to 55 percent (Uganda Bureau of Statistics, 2007, 2017).[59] In addition, the volume of health-related programming in rural Uganda increased significantly over that time span. Appendix Table G1 compares the levels of several key outcome measures at baseline in P2P and ACT Health and reveals significant differences. Given the magnitude of these differences, we cannot rule out the possibility that differing baseline conditions may be responsible for the differences we see across the two studies. It may simply be easier to improve health outcomes when they are starting at a lower baseline level or when the health intervention space is less crowded.[60] Highly suggestive validation of this possibility is provided in Appendix E.10, where we re-run our analysis in the sub-sample of health centers whose baseline child mortality rates are within one standard deviation of those reported at baseline in P2P and find significant treatment effects on child mortality (although no impact on utilization, treatment quality or other health outcomes). These results are robust to different specifications and a correction for multiple comparisons (see Appendix E.10).[61]

8 Conclusion

One of the most influential ideas in development circles in recent years is that service delivery shortfalls in low-income countries stem from an "information gap" that prevents citizens from holding local service providers accountable for poor performance (Khemani, 2007; Mansuri and Rao, 2013; Kosack and Fung, 2014; World Bank, 2016). The presumed solution to this problem is to provide citizens with information—often in the form of a report card—about the quality of services being provided to their community, sometimes along with information about government standards that are supposed to be (but are often not) met and/or comparative information about the quality of service provision in other communities. Armed with this information, it is thought

[59] Figures from the Uganda Demographic and Health Surveys from 2006 and 2016, respectively, since they are not reported annually. Stunting and wasting are defined as height-for-age and weight-for-height, respectively, that are more than 2 standard deviations below the median of the WHO child growth standards.

[60] This may account for why, in preliminary results, Dube, Haushofer and Siddiqi (2018) find treatment impacts on utilization and child mortality (although not on service quality, patient satisfaction, or other health outcomes) in a P2P-inspired community monitoring health intervention in Sierra Leone: the under-five mortality rate in Sierra Leone at the time of their study was 182 per 1,000 live births—even higher than in Uganda at the time of P2P.

[61] We also re-estimated treatment impact in the even more restricted set of units (just 29 health centers) that matched P2P in both baseline child mortality rates and health center level. We find that exposure to treatment is associated with a statistically significant decline in child mortality, although only when calculated at the child-level. The point estimate on child mortality in the health center-level analysis is negative, but, due to the small sample size, is no longer statistically significant.

that citizens can then apply pressure on underperforming service providers, who will respond by improving their performance.

The findings of our evaluation of the ACT Health program—a large-scale community monitoring intervention modeled precisely on the received wisdom about the power of information to initiate this causal process—raise questions about the usefulness of this development strategy. While we find evidence for marginal effects of the intervention on treatment quality, we find no evidence for its impact on utilization or health outcomes (including child mortality)—the bottom-line outcomes that development practitioners ultimately want to affect. We also find no evidence that the intervention caused citizens to more closely monitor their local health care providers or to apply pressure on those who were revealed to be underperforming.

Our findings with respect to treatment quality suggest that providing information about service quality can indeed lead to (small) changes in the behavior of frontline service providers. But our findings with respect to community monitoring indicate that the link between information provision and provider behavior does not run through citizen pressure. Indeed, our findings regarding the impact of having government officials present during the intervention suggest that providing information to top-down principals may be a stronger lever for changing the behavior of frontline health providers than mobilizing pressure from the bottom-up. Moreover, our robust null result with respect to health outcomes (including child mortality) suggest that changes in the behavior of health providers do not necessarily lead to measurable improvements in the well being of the community that the health providers serve—at least for the health outcomes we measure and within the time frame we study.[62]

Our results underscore the difficulty in changing bottom-line health outcomes and suggest that a sharper distinction should be made between outcomes that may be directly affected by a development intervention, such as health provider behavior, and outcomes that are products of more complex causal processes in which the outcomes that the intervention can plausibly move may only play a minor role, such as utilization and child health.

Our study also has implications for the broader objective of knowledge accumulation in the social sciences. Notwithstanding our attempt to model ACT Health as closely as possible on P2P, our discussion of the many differences across the two studies with respect to study populations, program implementation, variable operationalization, econometric specifications, and baseline conditions underscores just how challenging it is to replicate (and, by doing so, attempt to confirm or disconfirm the findings of) a prior field experiment. For these reasons, the ACT Health should be seen not as a "replication" of P2P but as what Clemens (2017) terms a "robustness test," since it employs different code from that used in the original study and data gathered from both a different population and at a different time.

Whether one puts greater stock in the original study or the subsequent one depends less on a judgment that one study or the other is "right" than on an appraisal of each study's power to

[62]While we think it is unlikely, we cannot rule out that we would have found treatment effects on other health outcomes—for example, disease rates or disability affected life years—had we measured them. We also cannot dismiss the possibility that the process of behavioral change created by an intervention like ACT Health takes more than 20 months to generate discernible effects on health outcomes or child mortality. If this is the case, then we measured our outcomes too soon. We can, however, rule out with some confidence that we measured our outcomes too *late*, since our null findings are little changed when we examine midline rather than endline data.

detect treatment effects, an assessment of the appropriateness of the econometric specifications and variable operationalizations that were employed to evaluate program impact, and on the match between each study's sample and the population to which one seeks to apply the findings. With respect to the last factor, the ACT Health findings may be more germane for understanding the likely impact of bottom-up interventions in the world today, as very few low-income countries have health outcomes as poor as those found in Uganda at the time of P2P, ten years earlier.

References

Alcott, Hunt. 2015. "Site selection bias in program evaluation." *Quarterly Journal of Economics* 130(3):1117–1165.

Algan, Yann, Camille Hémet, David Laitin et al. 2016. "The social effects of ethnic diversity at the local level: A natural experiment with exogenous residential allocation." *Journal of Political Economy* 124(3):696–733.

Andrabi, Tahir, Jishnu Das and Asim Ijaz Khwaja. 2017. "Report cards: The impact of providing school and child test scores on educational markets." *American Economic Review* 107(6).

Arinaitwe, Solomon. 2017. "Drug thefts mar hospitals as patients bear brunt." Daily Monitor, December 3, 2017.
URL: *https://www.monitor.co.ug/SpecialReports/Drug-thefts-mar-hospitals-as-patients-bear-the-brunt/688342-4211180-2o6l4pz/index.html*

Arrow, Kenneth. 1974. "Essays in the theory of risk-bearing." *Journal of Business* 47(1).

Banerjee, Abhijit, Rema Hanna, Jordan Kyle, Benjamin A. Olken and Sudarno Sumarto. 2018. "Tangible information and citizen empowerment: Identification cards and food subsidy programs in Indonesia." *Journal of Political Economy* 126(2):451–491.

Banerjee, Abhijit, Rukmini Banerji, Esther Duflo, Rachel Glennerster and Stuti Khemani. 2010. "Pitfalls of participatory programs: Evidence from a randomized evaluation in education in India." *American Economic Journal: Economic Policy* 2(1):1–30.

Banerjee, Abhijit, Rukmini Banerji, James Berry, Esther Duflo, Harini Kannan, Shobhini Mukerji, Marc Shotland and Michael Walton. 2017. "From proof of concept to scalable policies: Challenges and solutions, with an application." *Journal of Economic Perspectives* 31(4):73–102.

Banerjee, Abhijit, Sylvian Chassang and Erik Snowberg. 2017. Decision theoretic approaches to experiment design and external validity. In *Handbook of economic field experiments*, ed. Abhijit Banerjee and Esther Duflo. pp. 141–174.

Barr, Abigail, Frederick Mugisha, Pieter Serneels and Andrew Zeitlin. 2012. "Information and collective action in community-based monitoring of schools: Field and lab experimental evidence from Uganda." Working paper.
URL: *https://www.dartmouth.edu/ neudc2012/docs/paper-277.pdf*

Beath, Andrew, Fotini Christia and Ruben Enikolopov. 2017. "Direct democracy and resource allocation: Experimental evidence from Afghanistan." *Journal of Development Economics* 124:199–213.

Benjamini, Yoav and Yosef Hochberg. 1995. "Controlling the false discovery rate: A practical and powerful approach to multiple testing." *Journal of the Royal Statistical Society. Series B (Methodological)* pp. 289–300.

Besley, Timothy. 2007. *Principled agents? The political economy of good government*. New York: Oxford University Press.

Björkman, Martina and Jakob Svensson. 2009. "Power to the people: Evidence from a randomized field experiment of a community-based monitoring project in Uganda." *Quarterly Journal of Economics* 124(2):735–769.

Björkman, Martina and Jakob Svensson. 2010. "When is community-based monitoring effective? Evidence from a randomized experiment in primary health in Uganda." *Journal of the European Economic Association* 8(2-3):571–581.

Björkman Nyqvist, Martina, Andrea Guariso, Jakob Svensson and David Yanagizawa-Drott. 2016. "Effect of a micro entrepreneur-based community health delivery program on under-five mortality in Uganda: A cluster-randomized controlled trial." CEPR Discussion Paper, DP11515.

Björkman Nyqvist, Martina, Damien de Walque and Jakob Svensson. 2017. "Experimental evidence on the long-run impact of community-based monitoring." *American Economic Journal: Applied Economics* 9(1):33–69.

Blair, Graeme, Jasper Cooper, Alexander Coppock and Macartan Humphreys. Forthcoming. "Declaring and diagnosing research designs." *American Political Science Review* .

BMAU. 2015. Continuous stock-outs of medical supplies in Uganda: What are the root causes? Technical report. Budget Monitoring and Accountability Unit (BMAU) Briefing Paper 15/15, Ministry of Finance, Planning, and Economic Development.

Bold, Tessa, Mwangi Kimenyi, Germano Mabu, Alice Ng'ang'a and Justin Sandefur. 2013. "Scaling up what works: Experimental evidence on external validity in Kenyan education." Center for Global Development Working Paper no. 321.

Brown, Annette N. and Benjamin D.K. Wood. 2019. "Replication studies of development impact evaluations." *Journal of Development Studies* 55(5):917–925.

Clemens, Michael A. 2017. "The meaning of failed replications: A review and proposal." *Journal of Economic Surveys* 31(1):326–342.

Cox, D. R. 1972. "Regression Models and Life-Tables." *Journal of the Royal Statistical Society* 34(2):187–220.

Donato, Katherine and Adrian Garcia Mosqueira. 2019. "Information improves provider behaviour: A replication study of a community-based monitoring programme in Uganda." *Journal of Development Studies* 55(5):967–988.

Dube, Oindrila, Johannes Haushofer and Bilal Siddiqi. 2018. "Top-down or bottom-up? Impacts of community monitoring and government awards on health care in Sierra Leone." Working Paper.

Dunning, Thad, Guy Grossman, Macartan Humphreys, Susan D. Hyde, and Craig McIntosh. Forthcoming. *Metaketa 1: The limits of electoral accountability*. Cambridge University Press.

Fearon, James D, Macartan Humphreys and Jeremy M Weinstein. 2015. "How does development assistance affect collective action capacity? Results from a field experiment in post-conflict Liberia." *American Political Science Review* 109(3):450–469.

Fiala, Nathan and Patrick Premand. 2018. "Social accountability and service delivery: Evidence from a large-scale experiment in Uganda." Policy Research Working Paper, WPS8449.

Grossman, Guy and Kristin Michelitch. 2018. "Information dissemination, competitive pressure, and politician performance between elections: A field experiment in Uganda." *American Political Science Review* 112(2).

Grossman, Guy, Macartan Humphreys and Gabriella Sacramone-Lutz. 2016. "Information technology and political engagement: Mixed evidence from Uganda." Working paper.
URL: *https://cpb-us-w2.wpmucdn.com/web.sas.upenn.edu/dist/7/228/files/2016/09/GHS-20190110-2aw7rnr.pdf*

Holmström, Bengt. 1979. "Moral hazard and observability." *The Bell Journal of Economics* pp. 74–91.

Kahn, Steven A., James C. Iannuzzi, Nicole A. Stassen, Paul E. Bankey and Mark Gestring. 2015. "Measuring satisfaction: factors that drive hospital consumer assessment of healthcare providers and systems survey responses in a trauma and acute care surgery population." *The American Surgeon* 81(5):537–543.

Keefer, Philip and Stuti Khemani. 2014. "Mass media and public education: The effects of access to community radio in Benin." *Journal of Development Economics* 109:57–72.

Khemani, Stuti. 2007. "Can information campaigns overcome political obstacles to serving the poor?" *The Politics of Service Delivery in Democracies: Better Access for the Poor* pp. 56–69.

Khwaja, Asim Ijaz. 2009. "Can good projects succeed in bad communities?" *Journal of Public Economics* 93(7):899–916.

Kling, Jeffrey R., Jeffrey B. Liebman and Lawrence F. Katz. 2007. "Experimental analysis of neighborhood effects." *Econometrica* 75(1):83–119.

Kosack, Stephen and Archon Fung. 2014. "Does transparency improve governance?" *Annual Review of Political Science* 17:65–87.

Lieberman, Evan S., Daniel N. Posner and Lily L. Tsai. 2014. "Does information lead to more active citizenship? Evidence from an education intervention in rural Kenya." *World Development* 60:69–83.

Lin, Winston, Donald Green and Alexander Coppock. 2016. "Standard operating procedures: A Safety net for pre-analysis plans." *PS: Political Science & Politics* 49(3):495–499.

Lin, Winston et al. 2013. "Agnostic notes on regression adjustments to experimental data: Reexamining Freedman's critique." *The Annals of Applied Statistics* 7(1):295–318.

Mansuri, Ghazala and Vijayendra Rao. 2013. *Localizing development: Does participation work?* Washington, DC: World Bank.

Miguel, Edward and Mary Kay Gugerty. 2005. "Ethnic diversity, social sanctions, and public goods in Kenya." *Journal of Public Economics* 89(11):2325–2368.

Molina, Ezequiel, Laura Carella, Ana Pacheco, Guillermo Cruces and Leonardo Gasparini. 2016. "Community monitoring interventions to curb corruption and increase access and quality of service delivery in low-and middle-income countries: a systematic review." *Campbell Systematic Reviews* 12.

Olken, Benjamin. 2007. "Monitoring corruption: Evidence from a field experiment in Indonesia." *Journal of Political Economy* 115(2):200–249.

Olken, Benjamin. 2010. "Political institutions and local public goods: Evidence from a field experiment in Indonesia." *American Political Science Review* 104(7):243–267.

Pandey, Priyanka, Sangeeta Goyal and Venkatesh Sundararaman. 2009. "Community participation in public schools: Impact of information campaigns in three Indian states." *Education Economics* 17(3):355–75.

Pradhan, Menno, Daniel Suryadarma, Amanda Beatty, Maisy Wong, Arya Gaduh, Arminda Alisjahbana and Rima Prama Artha. 2009. "Improving educational quality through enhancing community participation: Results from a randomized field experiment in Indonesia." *American Economic Journal: Applied Economics* 6(2):105–126.

Raffler, Pia. 2017. "Does political oversight of the bureaucracy increase accountability? Field experimental evidence from an electoral autocracy." Working paper.
URL: *http://piaraffler.com/wp-content/uploads/2018/09/Raffler-Political-Oversight.pdf*

Rosner, Bernard. 2015. *Fundamentals of Biostatistics*. Nelson Education.

Ross, Stephen A. 1973. "The economic theory of agency: The principal's problem." *The American Economic Review* 63(2):134–139.

Rowland, Donald T. 2003. *Demographic methods and concepts*. New York: Oxford University Press.

Rutaremwa, Gideon, Stephen Ojiambo Wandera, Tapiwa Jhamba, Edith Akiror and Angela Kiconco. 2015. "Determinants of maternal health service utilization in Uganda." *BMC Health Services Research* 15:271.

Rutstein, Shea Oscar and Guillermo Rojas. 2006. *Guide to DHS statistics*. Calverton, MD: ORC Macro.

Rwothungeyo, Billy. 2016. "Procurement reforms: Has health service delivery improved?" New Vision, June 1 2016.
URL: *https://www.newvision.co.ug/new-vision/news/1425875/procurement-reforms-health-service-delivery-improved*

Tsai, Lily L. 2007. "Solidary groups, informal accountability, and local public goods provision in rural China." *American Political Science Review* 101(2):355–372.

Uganda Bureau of Statistics. 2007. *Uganda Demographic and Health Survey, 2006*. Kampala, Uganda: Bureau of Statistics.

Uganda Bureau of Statistics. 2017. *Uganda Demographic and Health Survey 2016*. Kampala, Uganda: Bureau of Statistics.

Uganda Ministry of Health. 2017. *Annual health sector performance report, 2016-2017*. Kampala, Uganda: Ministry of Health.

UNDG. 2003. *Indicators for monitoring the Millennium Development Goals: definitions, rationale, concepts and sources*. Vol. 95 United Nations Publications.

UNICEF. 2017. *Levels and trends in child mortality*. New York: United Nations.

World Bank. 2004. *World Development Report 2004: Making services work for poor people*. Washington, DC: World Bank.

World Bank. 2016. *Making politics work for development: Harnessing transparency and citizen engagement*. Washington, DC: World Bank.

Zwane, Alix Peterson, Jonathan Zinman, Eric Van Dusen, William Pariente, Clair Null, Edward Miguel, Michael Kremer, Dean S Karlan, Richard Hornbeck, Xavier Giné et al. 2011. "Being surveyed can change later behavior and related parameter estimates." *Proceedings of the National Academy of Sciences* 108(5):1821–1826.

Appendices

A Intermediate Outcomes

Table A1: Intermediate outcome indices and their components

	Mean
Citizen knowledge	
Share of patients' rights that household head is able to name correctly	10.20%
Share of patients' responsibilities that household head is able to name correctly	30.12%
Share of services offered at health center that household head is able to name correctly	64.76%
Health center staff knowledge	
Share of patients' rights that health center staff is able to name correctly	31.97%
Share of patients' responsibilities that health center staff is able to name correctly	54.93%
Efficacy	
Whether household head thinks she has "a lot"/"some" power to improve quality of health care at local HC	33.98%
Whether household head thinks she would be able to pressure a health worker to exert better effort	62.79%
Whether household head thinks she would be able to pressure a health worker to report to work on time ◇	61.93%
Whether household head thinks she has "a lot"/"some" influence in making village a better place to live ◇	33.65%
Whether household head agrees that "people like you have a say about how the government provides health care to your community"	82.32%
Whether household head agrees that "people like you have a say about how health facilities provide health care to your community"	81.18%
Community responsibility	
Whether household head thinks she is responsible for making sure health workers come to work and provide high quality health services	45.15%
Whether household head thinks community members are responsible for making sure health workers come to work and provide high quality health services	1.24%
Community monitoring	
Whether household members report having attended LC1 meetings in the last year	88.93%
Whether household members who attended LC1 meeting report that local health center was discussed	65.93%
Whether household members think engaged community members would find out if a health worker did not provide the effort that he/she should in caring for his/her patients ◇	73.04%
Whether household members think engaged community members would find out if a health worker did not report for work ◇	73.52%
Relationship between health care workers and the community	
Whether household members report being "satisfied"/"very satisfied" with relationship with health center staff	73.67%
Whether household members say they trust the workers at the health center	60.10%
Whether health center staff report being "satisfied"/"very satisfied" with their relationship with the community	90.62%
Whether household members did not say that the health center staff would "refuse to see me" or "behave hostilely toward me" if they had a complaint about the quality of services at the health center and decided to talk to the facility staff	97.69%
Health center transparency	
Whether a poster showing health center's opening/closing hours was visible during unannounced visit	2.78%
Whether a staff duty roster was displayed publicly during unannounced visit	20.31%
Whether a suggestion box was present during unannounced visit	6.14%
Whether information was posted listing services provided at the health center during unannounced visit	33.14%
Whether information was posted about patients' rights and responsibilities during the announced visit	3.46%

◇ Baseline values for this variable were not collected; values shown are from the control group. The baseline index omits these components.

B Procedures to Ensure Data Quality

The IPA ACT Health project team followed a standard set of procedures and processes developed by IPA over the years to manage large-scale academic research projects. These protocols include specific requirements for adhering to human subjects' regulations, developing survey instruments, fielding data collection teams, implementing data quality reviews, and producing and cleaning datasets for analysis.

The ACT Health project submitted research protocols for the three waves of data collection to both a local IRB committee (Mildmay/MUREC and UNCST) and to IPA's own internal IRB review committee (#2127). The project team worked closely with local authorities and received approvals for its work from the Office of the President, the Ugandan Ministry of Health, and from the chief administrative officers and district health officers in each of the 16 districts in which the project was implemented.

All personnel who handled the data and identified surveys in the field were required to obtain IRB training certificates. All field officers (including surveyors) signed confidentiality forms and administered informed consent to every respondent.

Data collection was undertaken in four distinct steps by teams headed by a research associate and consisting of field managers, team leaders, enumerators (health center, household, and anthropometric), mobilizers, trackers, and auditors.

1. Mobilization: A team of trained mobilizers contacted targeted households a day prior to the start of data collection to alert them to the survey work to come and to document the locations in which surveys would be administered. The conditions of the studied health center was assessed, staff attendance was recorded, and drug supplies were checked during a surprise visit to the health center on the same day. In-charges were notified to prepare the relevant records for the enumeration team's visit the following day.

2. Enumeration: The enumeration team completed four different surveys.

    (a) Household Survey: Household enumerators completed an average three to four surveys a day. The household survey took about one and a half hours to complete. Enumerators were instructed to interview the female head of the household. In the event that the female household head was no longer living in the house (at midline or endline), enumerators were instructed to follow the decision tree below to interview the correct person. During the interview, enumerators were asked to assess the number of children under five present in the household and to complete a form that household members would later give to anthropometric enumerators during their visits. At the end of the interview, contact forms were given to the household with instructions on how to report any comments to the HR management or IRB committee.

    (b) Anthropometric Survey: Anthropometric enumerators were specially trained to measure the height, weight and middle-upper arm circumference of all children in the household under 5 years old. Anthropometric enumerators were in charge of collecting the form left by household enumerators to ensure that the household survey was administered.

(c) Health Center Survey: Each health center enumerator (three per team) completed one survey a day. The survey consisted of interviewing the in-charges (at endline, a survey of other staff members was also added), assessing the quality of the health center, and collecting administrative reports (HMIS, PHC funds, etc.). Health center enumerators' visits were announced but could not take place the day of an immunization campaign, when health center staff were occupied. Health center enumerators were instructed to take pictures of administrative report pages with their PDA to prevent misreporting. These pictures were deleted each evening by the field team leaders.

(d) LC1 Survey: Team Leaders were responsible for the LC1 surveys, which involved interviewing the LC1 chairman about the characteristic of the village (rural/urban), its social cohesion, the political affiliations of officials, and other topics.

3. Tracking: Household that could not be found on the day of the enumeration were tracked by a team of trackers who were also trained to do the anthropometric survey. Tracking sheets were given to trackers by field managers after receiving approval from the research associate, following the decision tree below.

4. Auditing: Auditors performed back checks and spot checks (with field managers) on daily basis. Auditors received auditing sheets from the research associate once household data collection was finalized. They reported the findings of their investigations and handled their surveys directly to the research associate.

The IPA ACT Health survey team followed a set of standard operating procedures to ensure high quality data collection. These included:

- High quality training for everyone involved in the data collection: A total of four different teams of 75 enumerators worked in the 16 different districts. Mobilizers, enumerators, auditors and team leaders went through one-week trainings before being selected, including a soft launch to put in practice what they had learned.

- High-frequency checks: Specific survey questions that were susceptible to typos or incoherence were audited every evening by the research associates and field managers to ensure data quality. Daily feedback was provided to enumerators based on the findings from the monitoring, back checks, and high-frequency checks.

- Back checks: During the survey itself, data auditors re-surveyed a random sub-sample of survey participants (on a portion of the survey) to monitor enumerators' performance and to confirm that enumerators were interviewing the correct respondents. Field managers monitored their teams and accompanied each enumerator at least once every week.

- Monitoring and supervision of data collection: Research associates traveled with the survey team throughout the five months of the data collection and across the 16 different districts to supervise the data collection process. Field managers were present in the villages in which enumeration was ongoing on daily basis in order to monitor the data collection and perform spot checks on randomly selected enumerators. Principal investigators were updated in weekly calls about the data collection and were consulted to solve problems as they arose.

C Attrition, Balance, and Spillover

Table C1: Attrition Across Treatment Arms

	(1) Full treatment	(2) Information treatment	(3) Interface treatment	(4) Control	(5) P-value difference (1)-(4)	(6) P-value difference (1)&(2)-(3)&(4)	(7) P-value difference (1)&(3)-(2)&(4)
Reinterview rate, baseline to midline	0.947	0.940	0.950	0.944	0.68	0.44	0.16
Reinterview rate, baseline to endline	0.956	0.949	0.958	0.957	0.88	0.29	0.40

Notes. Numbers reported correspond to the average of household participation at health center level for the four different arms. The unit of observation is the health center catchment area. Columns (5)-(7) report the p-values of two-sided t-tests comparing re-interview rates in columns (1) and (4), (2) and (4), and (3) and (4), respectively. The full sample is composed of 376 health centers. At baseline, 379 health centers were surveyed but 3 dropped of the selected sample due to external factors (moved to another location, closed due to district decision or structural damages from flood).

Table C2: Spillover

	Mean close control HC	Mean far control HC	P-value difference
A. Difference Midline - Baseline levels of key outcome indices			
Utilization	0.006	-0.011	0.652
Treatment quality	-0.082	0.171	0.000
Patient satisfaction	0.012	-0.021	0.525
Health outcomes (HH level averaged at HC level)	-0.070	-0.034	0.621
Health outcomes (Child level averaged at HC level)	0.009	0.010	0.988
Child Mortality	0.010	0.029	0.286
B. Difference Endline - Baseline levels of key outcome indices			
Utilization	-0.009	-0.008	0.995
Treatment quality	-0.049	0.111	0.008
Patient satisfaction	-0.002	0.004	0.899
Health outcomes (HH level averaged at HC level)	0.033	0.008	0.714
Health outcomes (Child level averaged at HC level)	0.045	-0.074	0.430
Child Mortality	0.011	0.049	0.023
N	64	31	

Notes. Difference in means test comparing changes in main outcomes among control health centers that are relatively close to a treatment health center and those that are relatively far away. *Close* indicates that the distance to the nearest treatment health center is below the 67th percentile, *far* indicates that it is above. The dependent variable is defined as the change from baseline to midline (panel A) and the change from baseline to endline (panel B), respectively.

Table C3: Balance Across Treatment Arms

	(1) Full treatment	(2) Information only	(3) Interface only	(4) Control	(5) P-value difference (1) - (4)	(6) P-value difference (1) & (2) - (3) & (4)	(7) P-value difference (1) & (3) - (2) & (4)
A. Characteristics of catchment area							
Avg. distance of households to HC	0.98	0.97	0.85	1.34	0.28	0.52	0.17
Avg. household wealth in catchment area	0.01	-0.02	-0.00	0.02	0.86	0.79	1.00
Log pop. density in 3km radius around HC	8.46	8.62	8.64	8.64	0.23	0.34	0.45
Avg. level of education of household head	7.59	7.45	7.43	7.66	0.73	0.86	0.73
Share of households that received a VHT visit in the last 12 months	0.47	0.48	0.48	0.48	0.82	0.76	0.98
Share of households that declared NGOs activity in their village	0.15	0.13	0.14	0.15	0.87	0.80	0.63
B. Characteristics of HC							
Share providing delivery services	0.64	0.54	0.62	0.66	0.76	0.34	0.61
Share having staff houses	0.82	0.80	0.78	0.82	0.92	0.85	0.73
# of trained medical staff	6.26	6.45	6.34	7.17	0.18	0.40	0.28
Share having piped water	0.14	0.12	0.13	0.15	0.91	0.77	0.91
Share having electricity (grid or solar)	0.75	0.67	0.71	0.75	0.97	0.71	0.68
Avg. distance to nearest other government HC in district	3.90	3.63	3.97	4.09	0.49	0.17	0.70
C. Baseline levels of key outcome indices							
Utilization	5.24	5.27	5.52	5.43	0.41	0.17	0.86
Treatment quality	12.23	12.68	12.11	11.93	0.59	0.29	0.76
Patient satisfaction	0.78	0.79	0.78	0.78	0.61	0.82	0.35
Health outcomes (HH level)	1.07	1.12	1.07	1.05	0.38	0.10	0.58
Health outcomes (Child level)	2.01	2.04	2.07	1.97	0.61	0.96	0.53
Child Mortality	0.05	0.05	0.04	0.04	0.26	0.13	0.88
D. Baseline levels of intermediate outcome indices							
Citizen knowledge	0.20	0.20	0.21	0.20	0.54	0.14	0.55
Health worker knowledge	0.45	0.44	0.42	0.43	0.39	0.26	0.98
Efficacy	0.58	0.56	0.57	0.59	0.11	0.09	0.61
Community responsibility	0.23	0.23	0.23	0.24	0.24	0.23	0.68
Community monitoring	0.79	0.77	0.78	0.77	0.23	0.93	0.10
Relationship between health workers and community	0.80	0.80	0.80	0.82	0.16	0.24	0.47
Health center transparency	0.14	0.13	0.12	0.14	0.88	0.52	0.66
N	92	92	97	95			

D Measuring Child Mortality

We use three different approaches to measure child mortality rates: a *synthetic cohort approach* to calculate mortality rates per health center catchment area, which is similar to the method used in the Demographic and Health Surveys (DHS), a *child-level indicator* for whether a specific child is alive or dead in a given month, and the share of children who died in a catchment area, which mirrors the *vital statistics approach* used in Björkman and Svensson (2009). We describe each in turn.

Originally, we planned to use a vital statistics approach to measure mortality rates, since this was the main approach used in Björkman and Svensson (2009).[63] The vital statistics method uses a simple ratio of deaths under a certain age to live births during a recall period (UNDG, 2003). However, we updated our pre-analysis plan to prioritize the synthetic cohort life table approach because it offered a more precise measure of mortality. The difference in the data required for each method is that the synthetic life table approach requires the dates (month and year) of birth and death for every child that died during the recall period. In contrast, the vital statistics approach only requires asking if any child under the age of five had died in the last 12 months and the age they were when they died.

At endline, we collected the month of birth and, if applicable, death, also retrospectively for all children recorded during baseline and midline. Since the birth and death of children in the family is a very salient event, we are in this case not concerned about recall bias. To the contrary, the second, retrospective round of data collection of the month and age of death proved to be a helpful verification exercise, during which it became evident that a considerable share of the children that had been reported as having died in the past 12 months during baseline or midline had in fact died much earlier.

With this life table data, we are also able to use an even more nuanced measure of child mortality *at the child level*. Since we have the month of birth and, if applicable, death, for all 20,598 children in our sampled households who were ever under the age of five or unborn at baseline and still lived in the household (if alive) at either midline or endline, we are able to create a panel dataset indicating whether each child is dead or alive in a given month over the course of the 36 month study period.[64] This dataset, in turn, allows us to run child-level survival analyses using a Cox proportional hazards model (Cox, 1972), an estimation approach widely used for the analysis of survival rates (Rosner, 2015). We show results from this approach in the appendix.

Synthetic cohort life table

The *synthetic cohort life table* approach is similar to the method used in the Demographic and Health Surveys (DHS) as described in Rutstein and Rojas (2006) and Rowland (2003).[65] The approach calculates the probability of dying before a certain age (expressed per 1,000 births) by dividing the total number of deaths under that age by the total number of child years of exposure to the risk of dying.

[63] Björkman and Svensson (2009) also use an alternative measure, a binary indicator for child death during the recall period.

[64] Twelve month recall period prior to the baseline, 12 months between baseline and midline, and 12 months between midline and endline.

[65] The approach has also been used to measure child mortality rates in a randomized evaluation of a community health promoter program in Uganda implemented by Living Goods and BRAC (Björkman Nyqvist et al., 2016).

Calculating the mortality rate using this approach involved the following steps:

- Step 1: Create an age variable for each month of the recall period between 12 months prior to the baseline and the endline survey. Our baseline survey was conducted between August and December, 2014 and the endline was conducted between August and December of 2016. We therefore created a variable for the age in months beginning in August 2013: $age_{aug13}; age_{sep13}; \ldots; age_{dec13}; \ldots; age_{dec16}$.
- Step 2: Create a binary variable for the month that a child died: $died_{aug13}; \ldots; died_{dec16}$.
- Step 3: Create an age at death variable for each month: $agedied_{aug13}; \ldots; agedied_{dec16}$.
- Step 4: For each month, find the total number alive at each age, 0-59 months.
- Step 5: Sum the total number for each age across all months. (Use the mean number for each age of the current and previous months.)
- Step 6: For each month, find the total number that died at each age.
- Step 7: Sum the total number that died at each age across all months.
- Step 8: Calculate the mortality rate for each age in months (number of deaths at age A/total number of months at age A).
- Step 9: Calculate the survival rate for each age in months (1-mortality rate).
- Step 10: Calculate the overall survival rate by multiplying the individual age-specific survival rates across the relevant ages (i.e. 0-59 months for under-five survival rate).
- Step 11: Find the overall mortality rate (1-survival rate).

Vital statistics approach

The vital statistics approach calculates child mortality as the ratio of dead children in a given age bracket over all children in a given age bracket, dead or alive, calculated per catchment area.

Child-level indicator

A 36-month panel dataset for 20,598 children (some of whom are born into the panel at a later stage or age out of it), indicating for each month whether a child is dead or alive.

E Supporting Tables

E.1 Main Outcomes

The following tables show regression results for the main outcome indices (as summarized in Figure 3) and their components. The index components are shown first in standardized and then in non-standardized forms.

Table E1: Main outcomes: Averaged z-score indices

	(1) **Utilization**	(2) **Treatment quality**	(3) **Patient satisfaction**	(4) **Health outcomes**	(5) **Child mortality**
Full treatment	0.027	0.070***	0.077***	-0.003	-0.011
	(0.022)	(0.026)	(0.024)	(0.027)	(0.008)
Constant	-0.020	0.000	-0.002	-0.510***	0.061***
	(0.016)	(0.020)	(0.018)	(0.022)	(0.006)
N	7,288	7,288	7,288	4,930	187
R^2	0.230	0.102	0.043	0.112	0.197
P-value (Full treatment = 0)	0.213	0.008	0.001	0.900	0.188
Adjusted p-value (FT)	0.266	0.020	0.007	0.900	0.266

Notes. Estimates from Equation 1 comparing outcomes between the full treatment arm and the control group. The unit of observation in columns (1)-(3) is the household, in columns (4) it is the child, and in column (5) the health center catchment area. All models include district fixed effects and demeaned baseline covariates, as well as their interaction with the treatment indicator. Robust standard errors are clustered at health center level. The row *P-value (Full treatment = 0)* shows p-values for a Wald test of the null hypothesis that the coefficient on *full treatment* is equal to zero. Adjusted p-values (FT) refer to p-values of the coefficient on *Full treatment* which are adjusted using the Benjamini-Hochberg method. *** p<0.01; ** p<0.05; * p<0.10

Table E2: Utilization index – Subcomponents

	(1) **Utilization index**	(2) Vaccination rates, children<36 months	(3) % of visits to HC, vs. other providers	(4) Number of visits to HC
Full treatment	0.027	0.054	0.034	-0.001
	(0.022)	(0.035)	(0.032)	(0.027)
Constant	-0.020	-0.011	-0.001	-0.017
	(0.016)	(0.025)	(0.023)	(0.020)
N	7,288	4,212	7,288	7,288
R^2	0.230	0.057	0.178	0.284
Adjusted p-value (FT)	0.266	0.370	0.446	0.981

Notes. Estimates are derived from from Equation 1, comparing outcomes between the Full treatment arm and the Control. All models include district fixed effects as well as demeaned baseline covariates and their interaction with the treatment indicator. Robust standard errors are clustered at the health center level. The dependent variable in column (1) is the averaged z-score index of the outcomes presented in columns (2)-(4). The latter are z-scores of (2) vaccination rates of children under 36 months, (3) share of visits to the designated health center versus other providers, (4) number of visits to the designated health center. Adjusted p-values (FT) refer to p-values of the coefficient on Full Treatment which are adjusted using the Benjamini-Hochberg method. *** p<0.01; ** p<0.05; * p<0.10.

Table E3: Utilization index – Non-standardized subcomponents

	(1) **Utilization index**	(2) Vaccination rates, children<36 months	(3) % of visits to HC, vs. other providers	(4) Number of visits to HC
Full treatment	0.027	0.022	0.007	-0.008
	(0.022)	(0.014)	(0.007)	(0.342)
Constant	-0.020	0.739***	0.231***	9.128***
	(0.016)	(0.017)	(0.007)	(0.316)
N	7,288	4,212	7,288	7,288
R^2	0.230	0.057	0.178	0.284
Mean control group endline	-0.013	0.787	0.326	15.327
Mean control group baseline	-0.008	0.755	0.377	14.186

Notes. Estimates are derived from from Equation 1, comparing outcomes between the Full treatment arm and the Control. All models include district fixed effects as well as demeaned baseline covariates and their interaction with the treatment indicator. Robust standard errors are clustered at the health center level. The dependent variable in column (1) is the averaged z-score index of the outcomes presented in columns (2)-(4). The latter are (2) vaccination rates of children under 36 months, (3) share of visits to the designated health center versus other providers, (4) number of visits to the designated health center. *** p<0.01; ** p<0.05; * p<0.10.

Table E4: Treatment quality index – Subcomponents

	(1) **Treatment quality index**	(2) Used equipment	(3) Waiting time	(4) Examined by trained staff	(5) Privacy during exam	(6) Received test when needed	(7) Diagnosis explained clearly	(8) % staff presence	(9) Facility cleanliness	(10) Drug availability
Full treatment	0.070***	0.056	0.002	0.009	0.073*	0.036	0.058*	0.054	0.118	0.229**
	(0.026)	(0.036)	(0.048)	(0.035)	(0.039)	(0.040)	(0.032)	(0.133)	(0.140)	(0.111)
Constant	0.000	0.031	-0.030	-0.000	-0.000	0.010	0.005	0.010	0.012	0.016
	(0.020)	(0.029)	(0.033)	(0.026)	(0.035)	(0.032)	(0.025)	(0.107)	(0.112)	(0.092)
N	7,288	7,288	7,288	7,288	7,288	7,288	7,288	187	187	187
R^2	0.102	0.023	0.084	0.026	0.039	0.030	0.019	0.299	0.166	0.421
Adjusted p-value (FT)	0.020	0.269	0.960	0.901	0.224	0.604	0.224	0.878	0.604	0.224

Notes. Estimates are derived from from Equation 1, comparing outcomes between the Full treatment arm and the Control. All models include district fixed effects as well as demeaned baseline covariates and their interaction with the treatment indicator. Robust standard errors are clustered at the health center level. The dependent variable in column (1) is the averaged z-score index of the components presented in columns (2)-(10). The latter are z-scores of (2) whether household members reported that, during their most recent visit to the health center, equipment was used during examination, (3) waiting time consisting of the total amount of time spent by the household members waiting for the initial consultation and the examination; whether household members declared that, during their most recent visit to the health center, (4) they were examined by trained health care staff, (5) they had privacy during their examination, (6) lab tests were administered, (7) their diagnosis was clearly explained to them; (8) percent of staff in attendance during an unannounced visit to the health center, measured at the health center level, (9) condition of the clinic (cleanliness of floors and walls, whether the clinic smelled as observed during unannounced visit to health center), measured at the health center level, (10) share of months in which stock cards indicated availability of six key tracer drugs in the past three months, measured at the health center level. Adjusted p-values (FT) refer to p-values of the coefficient on Full Treatment which are adjusted using the Benjamini-Hochberg method. *** p<0.01; ** p<0.05; * p<0.10.

Table E5: Treatment quality index – Non-standardized subcomponents

	(1) **Treatment quality index**	(2) Used equipment	(3) Waiting time	(4) Examined by trained staff	(5) Privacy during exam	(6) Received test when needed	(7) Diagnosis explained clearly	(8) % staff presence	(9) Facility cleanliness	(10) Drug availability
Full treatment	0.070***	0.021	0.142	0.000	0.015*	0.013	0.023*	0.017	0.016	0.059**
	(0.026)	(0.013)	(2.829)	(0.002)	(0.008)	(0.014)	(0.013)	(0.041)	(0.019)	(0.029)
Constant	0.000	0.759***	69.736***	0.831***	0.910***	0.742***	0.697***	0.379***	0.691***	0.731***
	(0.020)	(0.014)	(2.228)	(0.152)	(0.012)	(0.015)	(0.013)	(0.043)	(0.056)	(0.099)
N	7,288	7,288	7,288	7,288	7,288	7,288	7,288	187	187	187
R^2	0.102	0.023	0.084	0.026	0.039	0.030	0.019	0.299	0.166	0.421

Notes. Estimates are derived from from Equation 1, comparing outcomes between the Full treatment arm and the Control. All models include district fixed effects as well as demeaned baseline covariates and their interaction with the treatment indicator. Robust standard errors are clustered at the health center level. The dependent variable in column (1) is the averaged z-score index of the components presented in columns (2)-(10). The latter are (2) whether household members reported that, during their most recent visit to the health center, equipment was used during examination, (3) waiting time consisting of the total amount of time spent by the household members waiting for the initial consultation and the examination; whether household members declared that, during their most recent visit to the health center, (4) they were examined by trained health care staff, (5) they had privacy during their examination, (6) lab tests were administered, (7) their diagnosis was clearly explained to them; (8) percent of staff in attendance during an unannounced visit to the health center, measured at the health center level, (9) condition of the clinic (cleanliness of floors and walls, whether the clinic smelled as observed during unannounced visit to health center), measured at the health center level, (10) share of months in which stock cards indicated availability of six key tracer drugs in the past three months, measured at the health center level. *** p<0.01; ** p<0.05; * p<0.10.

Table E6: Patient satisfaction index – Subcomponents

	(1) **Patient satisfaction index**	(2) Satisfied by HC quality	(3) Satisfied with quality of care	(4) Polite staff	(5) Staff interested in health	(6) Staff listening	(7) Free to express clearly	(8) Availability of staff improving
Full treatment	0.077***	0.105***	0.061*	0.074**	0.101***	0.071**	0.040	0.078*
	(0.024)	(0.035)	(0.032)	(0.035)	(0.030)	(0.031)	(0.039)	(0.040)
Constant	-0.002	-0.007	0.017	0.016	0.020	0.013	0.020	-0.024
	(0.018)	(0.023)	(0.023)	(0.027)	(0.024)	(0.025)	(0.029)	(0.032)
N	7,288	7,288	7,288	7,288	7,288	7,288	7,288	7,288
R^2	0.043	0.066	0.044	0.024	0.018	0.023	0.019	0.053
Adjusted p-value (FT)	0.007	0.010	0.065	0.060	0.006	0.055	0.310	0.065

Notes. Estimates are derived from from Equation 1, comparing outcomes between the Full treatment arm and the Control. All models include district fixed effects and demeaned baseline covariates, as well as their interaction with the treatment indicator. Robust standard errors are clustered at the health center level. The dependent variable in column (1) is the averaged z-score index of the outcomes presented in columns (2)-(8). The latter are z-scores of indicator variables of whether household members declared that (2) the services currently offered at the health center are of "very high quality" or "somewhat high quality", (3) they were "very satisfied" or "satisfied" with the quality of care received during their most recent visits to the health center, (4) during their most recent visit to the health center, the person conducting the examination behaved politely/showed respect, (5) during their most recent visit to the health center, the person conducting the examination appeared to be interested in their health condition, (6) during their most recent visit to the health center, the person conducting the examination listened to what they had to say, (7) during their most recent visit to the health center, they felt free to express themselves to the person conducting the examination, (8) compared to the year before, the availability of medical staff had improved at the health center. Adjusted p-values (FT) refer to p-values of the coefficient on Full Treatment which are adjusted using the Benjamini-Hochberg method. *** p<0.01; ** p<0.05; * p<0.10.

Table E7: Patient satisfaction index – Non-standardized subcomponents

	(1) **Patient satisfaction index**	(2) Satisfied by HC quality	(3) Satisfied with quality of care	(4) Polite staff	(5) Staff interested in health	(6) Staff listening	(7) Free to express clearly	(8) Availability of staff improving
Full treatment	0.077***	0.052***	0.023*	0.016**	0.023***	0.016**	0.011	0.035*
	(0.024)	(0.017)	(0.012)	(0.007)	(0.007)	(0.007)	(0.011)	(0.018)
Constant	-0.002	0.423***	0.712***	0.878***	0.883***	0.869***	0.840***	0.394***
	(0.018)	(0.011)	(0.012)	(0.014)	(0.014)	(0.014)	(0.014)	(0.015)
N	7,288	7,288	7,288	7,288	7,288	7,288	7,288	7,288
R^2	0.043	0.066	0.044	0.024	0.018	0.023	0.019	0.053

Notes. Estimates are derived from from Equation 1, comparing outcomes between the Full treatment arm and the Control. All models include district fixed effects and demeaned baseline covariates, as well as their interaction with the treatment indicator. Robust standard errors are clustered at the health center level. The dependent variable in column (1) is the averaged z-score index of the outcomes presented in columns (2)-(8). The latter are indicator variables of whether household members declared that (2) the services currently offered at the health center are of "very high quality" or "somewhat high quality", (3) they were "very satisfied" or "satisfied" with the quality of care received during their most recent visits to the health center, (4) during their most recent visit to the health center, the person conducting the examination behaved politely/showed respect, (5) during their most recent visit to the health center, the person conducting the examination appeared to be interested in their health condition, (6) during their most recent visit to the health center, the person conducting the examination listened to what they had to say, (7) during their most recent visit to the health center, they felt free to express themselves to the person conducting the examination, (8) compared to the year before, the availability of medical staff had improved at the health center. *** p<0.01; ** p<0.05; * p<0.10.

Table E8: Health outcomes index at the child level – Subcomponents

	(1) **Health outcomes index**	(2) **Weight/Age 0-18 months**	(3) **MUAC 0-18 months**	(4) **Weight/Age 18-36 months**	(5) **MUAC 18-36 months**
Full treatment	-0.003	-0.000	-0.015	0.004	0.017
	(0.027)	(0.048)	(0.047)	(0.031)	(0.028)
Constant	-0.510***	0.006	0.007	-0.463***	-0.640***
	(0.022)	(0.037)	(0.037)	(0.027)	(0.019)
N	4,930	2,140	2,140	2,790	2,790
R^2	0.112	0.018	0.018	0.225	0.346
Adjusted p-value (FT)	0.900	0.993	0.993	0.993	0.993

Notes. Estimates from Equation 1 comparing outcomes between the Full treatment arm and the Control. All models include district fixed effects and demeaned baseline covariates, as well as their interaction with the treatment indicator. Robust standard errors are clustered at the health center level. The dependent variable in column (1) is the averaged z-score index of the components presented in columns (2)-(7). The latter are z-scores of (2) the average ratio of weight over number of months for children under 18 months, (3) the average ratio of weight over number of months for children 18-36 months old, (4) the average ratio of upper arm circumference over number of months for children under 18 months, (5) the average ratio of upper arm circumference over number of months for children 18-36 months old. The unit of analysis is the child. Adjusted p-values (FT) refer to p-values of the coefficient on Full Treatment which are adjusted using the Benjamini-Hochberg method. *** p<0.01; ** p<0.05; * p<0.10

Table E9: Health outcomes index at the child level – Non-standardized subcomponents

	(1) **Health outcome overall**	(2) **Weight/Age 0-18 months**	(3) **MUAC 0-18 months**	(4) **Weight/Age 18-36 months**	(5) **MUAC 18-36 months**
Full treatment	-0.003	-0.000	-0.041	0.000	0.002
	(0.027)	(0.048)	(0.132)	(0.002)	(0.003)
Constant	-0.510***	1.351***	2.874***	0.363***	0.452***
	(0.022)	(0.139)	(0.327)	(0.003)	(0.003)
N	4,930	2,140	2,140	2,790	2,790
R^2	0.112	0.018	0.018	0.225	0.346

Notes. Estimates from Equation 1 comparing outcomes between the Full treatment arm and the Control. All models include district fixed effects and demeaned baseline covariates, as well as their interaction with the treatment indicator. Robust standard errors are clustered at the health center level. The dependent variable in column (1) is the averaged z-score index of the components presented in columns (2)-(7). The latter are (2) the average ratio of weight over number of months for children under 18 months, (3) the average ratio of weight over number of months for children 18-36 months old, (4) the average ratio of upper arm circumference over number of months for children under 18 months, (5) the average ratio of upper arm circumference over number of months for children 18-36 months old. The unit of analysis is the child. *** p<0.01; ** p<0.05; * p<0.10

Table E10: Child mortality at the HC level

	(1) Child mortality 0-5 years old	(2) Child mortality 0-1 year old	(3) Child mortality 1-5 years old
Full treatment	-0.011	-0.006	-0.005
	(0.008)	(0.007)	(0.004)
Constant	0.061***	0.041***	0.020***
	(0.006)	(0.005)	(0.003)
N	187	187	187
R^2	0.197	0.211	0.184
Adjusted p-value (FT)	0.266	0.383	0.383

Notes. Estimates from equation 1 comparing the full treatment to the control group. The unit of observation is health center catchment area. The dependent variable is the mortality rate calculated using a synthetic cohort approach for the age brackets 0-5 years (1), 0-12 months (2), and 1-5 years (3), respectively. Adjusted p-values (FT) refer to p-values of the coefficient on Full Treatment which are adjusted using the Benjamini-Hochberg method. *** p<0.01; ** p<0.05; * p<0.10

E.2 Intermediate Outcomes

The following tables show regression results for the seven intermediate outcome indices (as summarized in Figure 5) and their components.

Table E11: Intermediate outcomes – Averaged z-score indices

	(1) **Citizen knowledge**	(2) **HC staff knowledge**	(3) **Efficacy**	(4) **Community responsibility**	(5) **Community monitoring**	(6) **Relationship**	(7) **HC transparency**
Full treatment	-0.056**	0.171	-0.022	-0.012	0.006	0.040	0.007
	(0.023)	(0.121)	(0.023)	(0.020)	(0.028)	(0.039)	(0.076)
Constant	-0.009	-0.016	-0.002	-0.002	0.003	0.004	-0.006
	(0.016)	(0.080)	(0.015)	(0.014)	(0.019)	(0.028)	(0.053)
N	7,288	187	7,288	7,288	7,288	7,288	187
R^2	0.205	0.276	0.045	0.054	0.097	0.047	0.481
P-value (Full treatment = 0)	0.017	0.158	0.343	0.538	0.838	0.307	0.930
Adjusted p-value (FT)	0.121	0.552	0.601	0.754	0.930	0.601	0.930

Notes. Estimates from Equation 1 comparing outcomes between the Full treatment arm and the Control for intermediate outcome indices. All models include district fixed effects and demeaned baseline covariates, as well as their interaction with the treatment indicator. Robust standard errors are clustered at the health center level. The row *P-value (Full treatment = 0)* shows p-values for a Wald test of the null hypothesis that the coefficient on *full treatment* is equal to zero. Adjusted p-values (FT) refer to p-values of the coefficient on Full Treatment which are adjusted using the Benjamini-Hochberg method. *** p<0.01; ** p<0.05; * p<0.10

Table E12: Citizen knowledge index – Subcomponents

	(1) **Citizen knowledge index**	(2) # of patients rights correctly named	(3) # of patients resp. correctly named	(4) # of HC services correctly named
Full treatment	-0.056**	-0.094***	-0.118***	0.042
	(0.023)	(0.028)	(0.032)	(0.037)
Constant	-0.009	-0.009	-0.015	-0.003
	(0.016)	(0.019)	(0.024)	(0.023)
N	7,288	7,288	7,288	7,288
R^2	0.205	0.091	0.166	0.286
Adjusted p-value (FT)	0.121	0.001	0.001	0.246

Notes. Estimates from Equation 1 comparing outcomes between the Full treatment arm and the Control. All models include district fixed effects and demeaned baseline covariates, as well as their interaction with the treatment indicator. Robust standard errors are clustered at the health center level. The dependent variable in column (1) is the averaged z-score index of the dependent variables in columns (2)-(4). The dependent variable in column (2) is the z-score of the number of patient *rights*, listed in the patient's charter of the Ministry of Health, correctly named by community members, in column (3) is is the z-score of the number of patient *responsibilities*, listed in the patient's charter of the Ministry of Health, correctly named by households members, in column (4) it is the z-score of the number of health center services correctly named by community members. Adjusted p-values (FT) refer to p-values of the coefficient on Full Treatment which are adjusted using the Benjamini-Hochberg method. *** p<0.01; ** p<0.05; * p<0.10

Table E13: HC staff knowledge index – Subcomponents

	(1) **HC staff knowledge index**	(2) # of patients rights correctly named	(3) # of patients resp. correctly named
Full treatment	0.171	0.222	0.128
	(0.121)	(0.139)	(0.151)
Constant	-0.016	-0.019	-0.013
	(0.080)	(0.095)	(0.093)
N	187	187	187
R^2	0.276	0.290	0.211
Adjusted p-value (FT)	0.552	0.223	0.397

Notes. Estimates from Equation 1 comparing outcomes between the Full treatment arm and the Control. All models include district fixed effects and demeaned baseline covariates, as well as their interaction with the treatment indicator. Robust standard errors are clustered at the health center level. The dependent variable in column (1) is the averaged z-score index of the components presented in columns (2) and (3). The dependent variable in column (2) is the z-score of the number of patient *rights*, listed in the patient's charter of the Ministry of Health, correctly named by the health center in-charge, in column (3) it is the z-score of the number of patient *responsibilities*, listed in the patient's charter of the Ministry of Health, correctly named by the health center in-charge. Adjusted p-values (FT) refer to p-values of the coefficient on Full Treatment which are adjusted using the Benjamini-Hochberg method. *** p<0.01; ** p<0.05; * p<0.10

Table E14: Efficacy index – Subcomponents

	(1) **Efficacy index**	(2) Have power to improve HC services	(3) Can pressure health worker (effort)	(4) Can pressure health worker (timely)	(5) Can make village a better place to live	(6) Influence over gov. about health services	(7) Influence over HC about services provided
Full treatment	-0.022	-0.034	-0.014	-0.032	-0.040	-0.025	0.008
	(0.023)	(0.028)	(0.033)	(0.038)	(0.032)	(0.031)	(0.028)
Constant	-0.002	0.001	-0.009	-0.008	0.007	-0.001	0.001
	(0.015)	(0.021)	(0.023)	(0.026)	(0.023)	(0.020)	(0.018)
N	7,288	7,288	7,288	7,288	7,288	7,288	7,288
R^2	0.045	0.060	0.025	0.018	0.040	0.031	0.037
Adjusted p-value (FT)	0.601	0.632	0.787	0.632	0.632	0.632	0.787

Notes. Estimates from Equation 1 comparing outcomes between the Full treatment arm and the Control. All models include district fixed effects and demeaned baseline covariates, as well as their interaction with the treatment indicator. Robust standard errors are clustered at the health center level. The dependent variable in column (1) is the averaged z-score index of the components presented in columns (2)-(7). The remaining dependent variables are z-scores of the following indicator variables for whether community members think they: (2) have power to improve the quality of health care at the designated health facility, (3) they can pressure health worker to exert better effort in caring for patients by reporting them, (4) they can pressure health worker to work on time by reporting them, (5) they have influence in making the designated village a better place to live, (6) they have a say about how authorities provide health care to their community, (7) they have a say about how health facilities provide health care to their community. Adjusted p-values (FT) refer to p-values of the coefficient on Full Treatment which are adjusted using the Benjamini-Hochberg method. *** p<0.01; ** p<0.05; * p<0.10

Table E15: Community responsibility index – Subcomponents

	(1) **Community responsibility index**	(2) Community resp. for monitoring HC	(3) Community members also responsible
Full treatment	-0.012	0.014	-0.037
	(0.020)	(0.027)	(0.026)
Constant	-0.002	0.001	-0.006
	(0.014)	(0.019)	(0.018)
N	7,288	7,288	7,288
R^2	0.054	0.059	0.039
Adjusted p-value (FT)	0.754	0.614	0.303

Notes. Estimates from Equation 1 comparing outcomes between the Full treatment arm and the Control. All models include district fixed effects and demeaned baseline covariates, as well as their interaction with the treatment indicator. Robust standard errors are clustered at the health center level. The dependent variable in column (1) is the averaged z-score index of the components presented in columns (2) and (3). The dependent variable in column (2) is the z-score of a dummy variable indicating whether respondents think that they themselves are responsible for making sure that health workers come to work and provide high-quality health services, in column (3) it is the z-score of a dummy variable indicating whether respondents think *community members* are responsible for making sure that health workers come to work and provide high-quality health services. Adjusted p-values (FT) refer to p-values of the coefficient on Full Treatment which are adjusted using the Benjamini-Hochberg method. *** p<0.01; ** p<0.05; * p<0.10

Table E16: Community monitoring index – Subcomponents

	(1) **Community monitoring index**	(2) Attended LC1 meetings	(3) HC discussed at LC1 meetings	(4) Community would find out: staff late	(5) Community would find out: staff no effort
Full treatment	0.006	0.049	0.032	-0.008	-0.046*
	(0.028)	(0.052)	(0.070)	(0.032)	(0.027)
Constant	0.003	0.044	0.026	0.008	0.005
	(0.019)	(0.041)	(0.049)	(0.024)	(0.019)
N	7,288	7,288	7,288	7,288	7,288
R^2	0.097	0.105	0.086	0.035	0.026
Adjusted p-value (FT)	0.930	0.692	0.796	0.796	0.354

Notes. Estimates from Equation 1 comparing outcomes between the Full treatment arm and the Control. All models include district fixed effects and demeaned baseline covariates, as well as their interaction with the treatment indicator. Robust standard errors are clustered at the health center level. The dependent variable in column (1) is the averaged z-score index of the components presented in columns (2)-(7), which are z-scores of the following variables: (2) a dummy variable whether household members report having attended at least one LC1 meeting during the last 12 months; (3) a dummy variable whether the local health center was discussed at the most recent LC1 meeting; (4) a Likert-scale variable of whether the community would find out if a staff were regularly late or (5) extended no effort. Adjusted p-values (FT) refer to p-values of the coefficient on Full Treatment which are adjusted using the Benjamini-Hochberg method. *** p<0.01; ** p<0.05; * p<0.10

Table E17: Relationship index – Subcomponents

	(1) **Relationship index**	(2) Community satisfied with relations	(3) Trust health workers	(4) HC staff satisfied with relations	(5) Health workers will listen to complaints
Full treatment	0.040	0.060*	0.076**	-0.042	0.001
	(0.039)	(0.031)	(0.035)	(0.146)	(0.030)
Constant	0.004	-0.000	0.006	-0.002	0.000
	(0.028)	(0.021)	(0.024)	(0.102)	(0.024)
N	7,288	7,288	7,288	187	7,288
R^2	0.047	0.043	0.046	0.167	0.009
Adjusted p-value (FT)	0.601	0.112	0.112	0.969	0.969

Notes. Estimates from Equation 1 comparing outcomes between the Full treatment arm and the Control. All models include district fixed effects and demeaned baseline covariates, as well as their interaction with the treatment indicator. Robust standard errors are clustered at the health center level. The dependent variable in column (1) is the averaged z-score index of the components presented in columns (2) and (3). The latter are indicators variables of whether households think that (2) they are responsible for making sure that health workers come to work and provide high quality health services and (3) that health center staff would listen to their complaints and would not refuse to see them or behave hostilely. Adjusted p-values (FT) refer to p-values of the coefficient on Full Treatment which are adjusted using the Benjamini-Hochberg method. *** p<0.01; ** p<0.05; * p<0.10

Table E18: HC transparency index – Subcomponents

	(1) **HC transparency index**	(2) Poster with opening times	(3) Duty roster displayed	(4) Suggestion box	(5) Info on services provided	(6) Info on patient rights
Full treatment	0.007	0.126	-0.142	0.010	-0.093	0.125
	(0.076)	(0.139)	(0.145)	(0.124)	(0.142)	(0.121)
Constant	-0.006	-0.018	-0.014	0.053	0.002	0.081
	(0.053)	(0.093)	(0.114)	(0.099)	(0.101)	(0.091)
N	187	187	187	187	187	187
R^2	0.481	0.328	0.210	0.398	0.276	0.462
Adjusted p-value (FT)	0.930	0.610	0.610	0.937	0.643	0.610

Notes. Estimates from Equation 1 comparing outcomes between the Full treatment arm and the Control. All models include district fixed effects and demeaned baseline covariates, as well as their interaction with the treatment indicator. Robust standard errors are clustered at the health center level. The dependent variable in column (1) is the averaged z-score index of the components presented in columns (2)-(6). The remaining dependent variables are z-scores of the following indicator variables whether the designated health facility has: (2) a poster with opening times (3) a duty roster table displayed; (4) a suggestion box; (5) a list of services provided displayed; (6) patient's rights displayed. Adjusted p-values (FT) refer to p-values of the coefficient on Full Treatment which are adjusted using the Benjamini-Hochberg method. *** p<0.01; ** p<0.05; * p<0.10

E.3 Treatment Effects on Subgroups of Health Centers

The following table provides further details on the results discussed in Section 5.4. It shows estimated average treatment effects for subgroups of health centers. Each set of subgroup effects is derived from a separate regression, estimated using Equation 3. The table displays the coefficient on Treatment for the base category (for example, HC2) and the linear combination of the coefficient on Treatment and on the interaction between Treatment and an indicator variable describing the specific subset (for example, Treat + Treat * HC3 for HC3).

Table E19: Subgroup effects on main outcomes

		(1) **Utilization**	(2) **Treatment quality**	(3) **Patient satisfaction**	(4) **Health outcomes**	(5) **Child mortality**
Health center level	HC2	0.029	0.093***	0.061*	-0.024	-0.008
		(0.028)	(0.035)	(0.033)	(0.036)	(0.011)
	HC3	0.026	0.038	0.100***	0.024	-0.013
		(0.036)	(0.041)	(0.035)	(0.045)	(0.013)
Treatment quality	High	0.035	0.050*	0.051*	0.049	-0.018
		(0.029)	(0.028)	(0.028)	(0.038)	(0.012)
	Low	0.018	0.090***	0.100***	-0.054	-0.002
		(0.026)	(0.030)	(0.032)	(0.038)	(0.013)
Alternative health care options	High	0.011	0.044	0.109***	0.008	-0.015
		(0.029)	(0.037)	(0.036)	(0.037)	(0.012)
	Low	0.059*	0.096**	0.040	-0.035	-0.003
		(0.033)	(0.038)	(0.035)	(0.042)	(0.012)
Embeddedness of HC staff	High	0.039	0.072*	0.100***	-0.029	-0.008
		(0.030)	(0.037)	(0.036)	(0.040)	(0.013)
	Low	0.017	0.068*	0.052	0.021	-0.012
		(0.033)	(0.038)	(0.036)	(0.038)	(0.009)
Collective action potential	High	0.015	0.052	0.055*	-0.034	-0.012
		(0.029)	(0.034)	(0.029)	(0.037)	(0.011)
	Low	0.044	0.093**	0.103**	0.041	-0.007
		(0.033)	(0.043)	(0.041)	(0.045)	(0.012)
Community monitoring	High	0.015	0.032	0.082***	-0.008	-0.004
		(0.025)	(0.032)	(0.027)	(0.039)	(0.011)
	Low	0.041	0.115***	0.069**	0.002	-0.019
		(0.031)	(0.030)	(0.032)	(0.036)	(0.012)
Efficacy	High	0.056**	0.063**	0.098***	-0.062	-0.020*
		(0.025)	(0.028)	(0.028)	(0.040)	(0.011)
	Low	0.003	0.076***	0.062**	0.045	-0.001
		(0.028)	(0.028)	(0.028)	(0.038)	(0.012)
Health NGOs present in the village	Yes	0.048	0.024	0.112***	-0.043	-0.015
		(0.034)	(0.042)	(0.038)	(0.091)	(0.013)
	No	0.013	0.098***	0.048	-0.007	-0.008
		(0.032)	(0.036)	(0.035)	(0.043)	(0.010)
Avg. distance to HC	High	-0.004	0.060*	0.038	0.003	-0.006
		(0.030)	(0.034)	(0.032)	(0.038)	(0.013)
	Low	0.063*	0.077*	0.120***	-0.008	-0.015
		(0.032)	(0.042)	(0.039)	(0.042)	(0.010)
Catchment area	Urban	0.109**	-0.030	0.066	-0.009	0.001
		(0.042)	(0.053)	(0.051)	(0.065)	(0.018)
	Rural	0.010	0.108***	0.073**	-0.008	-0.012
		(0.025)	(0.030)	(0.029)	(0.031)	(0.009)

Notes. This table shows estimated average treatment effects for subgroups of health centers. Each set of subgroup effects is derived from a separate regression, estimated using Equation 3. The table displays the coefficient on Treatment for the base category (for example, HC2) and the linear combination of the coefficient on Treatment and on the interaction between Treatment and an indicator variable describing the specific subset (for example, Treat + Treat * HC3 for HC3). For continuous variables, *High* indicates that a health center's value for the given variable is at or above the median; *Low* indicates that it is below the median. All models include district fixed effects and demeaned baseline covariates, as well as their interaction with the treatment indicator. Robust standard errors are clustered at the health center level. *** p<0.01; ** p<0.05; * p<0.10

Table E20: Subgroup effects on intermediate outcomes

		(1) **Citizen knowledge**	(2) **HC staff knowledge**	(3) **Efficacy**	(4) **Community responsibility**	(5) **Community monitoring**	(6) **Relationship**	(7) **HC transparency**
Health center level	HC2	0	0	0	0	0	0	0
	HC3	-	0	0	0	0	0	0
Treatment quality	High	0	0	0	0	0	0	0
	Low	-	+	0	0	0	0	0
Alternative health care options	High	0	0	0	0	0	+	0
	Low	-	+	0	0	0	0	0
Embeddedness of HC staff	High	0	0	0	0	0	0	0
	Low	0	0	0	0	0	0	0
Collective action potential	High	-	0	0	0	0	0	-
	Low	0	+	0	0	0	+	+
Community monitoring	High	0	0	0	0	0	0	0
	Low	-	+	0	0	0	+	0
Efficacy	High	0	0	0	0	0	0	0
	Low	-	0	0	0	0	0	+
Health NGOs present in the village	Yes	0	+	0	0	0	+	0
	No	0	0	0	0	0	0	0
Avg. distance to HC	High	-	0	-	0	0	0	0
	Low	0	0	0	0	0	+	0
Catchment area	Urban	-	0	0	0	0	0	0
	Rural	0	0	0	0	0	0	0

Notes. This table shows estimated average treatment effects for subgroups of health centers. Each set of subgroup effects is derived from a separate regression, estimated using Equation 3. The table displays the coefficient on Treatment for the base category (such as for example HC2) and the linear combination of the coefficient on Treatment and on the interaction between Treatment and an indicator variable describing the specific subset (such as for example Treat + Treat * HC3 for HC3). For continuous variables, *High* indicates that a health center's value for the given variable is at or above the median, Low indicates that it is below the median. All models include district fixed effects and demeaned baseline covariates, as well as their interaction with the treatment indicator. Robust standard errors are clustered at the health center level. + indicates a positive coefficient significant at least at the .1 level, - indicates a negative coefficient significant at least at the .1 level, and 0 indicates an insignificant coefficient. A dark shade of gray indicates significance at the 99% level, medium gray at the 95% level, and light gray at the 90% level. p <0.01; p <0.05; p <0.10.

Table E21: Subgroup effects on intermediate outcomes

		(1) **Citizen knowledge**	(2) **HC staff knowledge**	(3) **Efficacy**	(4) **Community responsibility**	(5) **Community monitoring**	(6) **Relationship**	(7) **HC transparency**
Health center level	HC2	-0.010	0.202	-0.014	0.008	0.003	0.015	-0.048
		(0.034)	(0.151)	(0.030)	(0.027)	(0.038)	(0.053)	(0.095)
	HC3	-0.118***	0.147	-0.032	-0.039	0.009	0.077	0.079
		(0.036)	(0.208)	(0.036)	(0.032)	(0.044)	(0.060)	(0.164)
Treatment quality	High	-0.022	-0.049	-0.011	-0.010	-0.004	0.025	0.075
		(0.029)	(0.197)	(0.028)	(0.030)	(0.032)	(0.037)	(0.110)
	Low	-0.095***	0.329*	-0.036	-0.017	0.013	0.048	-0.031
		(0.029)	(0.167)	(0.029)	(0.027)	(0.035)	(0.049)	(0.115)
Alternative health care options	High	-0.025	0.015	-0.020	0.016	0.019	0.122**	0.036
		(0.045)	(0.207)	(0.032)	(0.028)	(0.041)	(0.052)	(0.120)
	Low	-0.082*	0.321*	-0.027	-0.042	-0.000	-0.045	0.025
		(0.043)	(0.169)	(0.037)	(0.033)	(0.044)	(0.057)	(0.129)
Embeddedness of HC staff	High	-0.057	0.236	-0.012	0.004	0.008	0.080	0.028
		(0.044)	(0.175)	(0.034)	(0.032)	(0.037)	(0.064)	(0.141)
	Low	-0.059	0.096	-0.030	-0.023	0.003	0.001	-0.033
		(0.046)	(0.181)	(0.032)	(0.028)	(0.042)	(0.049)	(0.110)
Collective action potential	High	-0.071*	-0.008	-0.035	0.017	-0.014	-0.046	-0.229**
		(0.040)	(0.158)	(0.030)	(0.028)	(0.035)	(0.054)	(0.115)
	Low	-0.045	0.382*	-0.004	-0.049	0.037	0.166***	0.320**
		(0.051)	(0.220)	(0.036)	(0.034)	(0.049)	(0.051)	(0.138)
Community monitoring	High	-0.043	0.022	-0.021	-0.023	-0.006	0.004	0.011
		(0.029)	(0.173)	(0.028)	(0.024)	(0.034)	(0.048)	(0.108)
	Low	-0.072**	0.332*	-0.024	-0.000	0.024	0.081*	0.004
		(0.030)	(0.192)	(0.032)	(0.030)	(0.037)	(0.046)	(0.120)
Efficacy	High	-0.034	0.012	-0.031	-0.019	0.008	0.053	-0.172
		(0.030)	(0.169)	(0.027)	(0.026)	(0.029)	(0.039)	(0.108)
	Low	-0.071**	0.322	-0.015	-0.003	0.007	0.029	0.182*
		(0.028)	(0.197)	(0.028)	(0.027)	(0.032)	(0.044)	(0.104)
Health NGOs present in the village	Yes	-0.067	0.343*	0.008	-0.023	0.043	0.107*	-0.097
		(0.047)	(0.191)	(0.036)	(0.035)	(0.041)	(0.056)	(0.137)
	No	-0.060	0.013	-0.046	-0.001	-0.026	-0.008	0.060
		(0.044)	(0.184)	(0.033)	(0.027)	(0.042)	(0.059)	(0.111)
Avg. distance to HC	High	-0.089**	0.265	-0.061*	-0.025	-0.002	-0.039	-0.044
		(0.043)	(0.173)	(0.035)	(0.028)	(0.039)	(0.059)	(0.118)
	Low	-0.029	0.026	0.023	0.008	0.019	0.129**	0.030
		(0.046)	(0.200)	(0.032)	(0.033)	(0.044)	(0.051)	(0.139)
Catchment area	Urban	-0.146**	-0.180	-0.059	-0.038	0.020	0.069	0.007
		(0.064)	(0.258)	(0.047)	(0.046)	(0.054)	(0.090)	(0.189)
	Rural	-0.038	0.262	-0.003	0.007	0.012	0.034	-0.023
		(0.036)	(0.161)	(0.028)	(0.024)	(0.035)	(0.045)	(0.109)

Notes. This table shows estimated average treatment effects for subgroups of health centers. Each set of subgroup effects is derived from a separate regression, estimated using Equation 3. The table displays the coefficient on Treatment for the base category (such as for example HC2) and the linear combination of the coefficient on Treatment and on the interaction between Treatment and an indicator variable describing the specific subset (such as for example Treat + Treat * HC3 for HC3). For continuous variables, *High* indicates that a health center's value for the given variable is at or above the median, Low indicates that it is below the median. All models include district fixed effects and demeaned baseline covariates, as well as their interaction with the treatment indicator. Robust standard errors are clustered at the health center level. *** p<0.01; ** p<0.05; * p<0.10

E.4 Robustness Checks for Endline Results

Table E22: Robustness check main outcomes

	(1) **Utilization**	(2) **Treatment quality**	(3) **Patient satisfaction**	(4) **Health outcomes**	(5) **Child mortality**
A: Without control variables					
Program impact	0.034	0.070***	0.075***	-0.014	-0.010
	(0.023)	(0.027)	(0.025)	(0.028)	(0.008)
Constant	-0.022	0.001	-0.001	-0.000	0.061***
	(0.017)	(0.021)	(0.018)	(0.022)	(0.006)
B: Without district fixed effects					
Program impact	0.013	0.064**	0.075***	-0.014	-0.009
	(0.028)	(0.029)	(0.025)	(0.029)	(0.008)
Constant	-0.012	0.003	-0.001	0.000	0.063***
	(0.020)	(0.023)	(0.018)	(0.023)	(0.006)
C: With outcome measures aggregated at HC level					
Program impact	0.035*	0.062**	0.080***	-0.012	-0.011
	(0.021)	(0.027)	(0.025)	(0.029)	(0.008)
Constant	-0.014	0.002	-0.001	-0.009	0.061***
	(0.014)	(0.021)	(0.019)	(0.022)	(0.006)
D: Difference between post and pre-treatment values					
Program impact	0.004	0.074***	0.067***	-0.023	-0.008
	(0.025)	(0.028)	(0.025)	(0.034)	(0.008)
Constant	-0.023*	0.020	-0.011	0.028	0.045***
	(0.012)	(0.016)	(0.015)	(0.040)	(0.005)
Observations (A & B)	7,288	7,288	7,288	4,930	187
Observations (C)	187	187	187	187	187
Observations (D)	14,576	14,576	14,576	9,860	374

Notes. Estimates from Equation 1 comparing outcomes between the Full treatment arm and the Control, with the following variations: Panel A shows results without covariates, panel B without district fixed effects, and panel C aggregates outcome measures and covariates at the health center level (the unit of randomization). Panel D shows results from a difference in difference estimation. *** p<0.01; ** p<0.05; * p<0.10

Table E23: Robustness checks – Intermediate outcomes

	(1) **Citizen knowledge**	(2) **HC staff knowledge**	(3) **Efficacy**	(4) **Community responsibility**	(5) **Community monitoring**	(6) **Relationship**	(7) **HC transparency**
A: Without control variables							
Program impact	-0.063*	0.156	-0.024	-0.012	0.006	0.042	-0.028
	(0.033)	(0.125)	(0.024)	(0.021)	(0.028)	(0.039)	(0.089)
Constant	-0.005	-0.012	-0.000	-0.002	0.003	0.002	0.004
	(0.025)	(0.084)	(0.016)	(0.014)	(0.019)	(0.028)	(0.059)
B: Without district fixed effects							
Program impact	-0.065	0.155	-0.024	-0.018	0.010	0.044	0.005
	(0.040)	(0.123)	(0.027)	(0.028)	(0.036)	(0.042)	(0.079)
Constant	-0.005	-0.009	-0.000	0.001	0.001	0.002	-0.007
	(0.028)	(0.081)	(0.017)	(0.021)	(0.026)	(0.030)	(0.057)
C: Without ouctome measures aggregated at HC level							
Program impact	-0.058**	0.171	-0.023	-0.013	-0.001	0.029	0.007
	(0.024)	(0.121)	(0.025)	(0.022)	(0.029)	(0.042)	(0.076)
Constant	-0.008	-0.016	-0.001	-0.001	0.004	0.002	-0.006
	(0.017)	(0.080)	(0.016)	(0.015)	(0.020)	(0.030)	(0.053)
D: Difference between post and pre-treatment values							
Program impact	-0.067*	0.188	-0.025	-0.017	0.013	0.030	-0.012
	(0.038)	(0.118)	(0.025)	(0.022)	(0.032)	(0.040)	(0.080)
Constant	-0.006	0.046	-0.018	-0.016	0.029	-0.042	0.003
	(0.020)	(0.055)	(0.011)	(0.010)	(0.021)	(0.026)	(0.042)
Observations (A & B)	7,288	187	7,288	7,288	7,288	7,288	187
Observations (C)	187	187	187	187	187	187	187
Observations (D)	14,576	374	14,576	14,576	14,576	14,576	374

Notes. Estimates from Equation 1 comparing outcomes between the Full treatment arm and the Control, with the following variations: Panel A shows results without covariates, panel B without district fixed effects, and panel C aggregates outcome measures and covariates at the health center level (the unit of randomization). Panel D shows results from a difference in difference estimation. *** p<0.01; ** p<0.05; * p<0.10

We also test the effect on child mortality with a Cox proportional hazards model, leveraging the fact that we have child-month level data on survival over the course of 36 months for over 20,000 children. By estimating the treatment effect on the chance of survival of the individual child, it mimics the data generation process more closely. The Cox model includes the same vector of controls and their interaction with the treatment indicator as Equation 1. Standard errors are clustered by health center catchment area.[66]

Table E24: Child mortality at the child level

	(1) **0-5 years old**	(2) **0-1 year old**	(3) **1-5 years old**
Full treatment	1.059	1.120	0.612
	(0.239)	(0.295)	(0.286)
N	10,118	4,543	8,635
Adjusted p-value (FT)	0.900	0.515	0.450

Notes. Displaying hazard ratios estimated with a Cox proportional hazards model, comparing outcomes between the full treatment and the control group. A hazard ratio below (above) 1 implies that the treatment led to lower (higher) mortality rates. All models include district fixed effects and demeaned baseline covariates, as well as their interaction with the treatment indicator. Standard errors are clustered at the health center level. The dependent variable is the incident of death, observed at the child-month level in the age bracket 0-5 years (1), 0-12 months (2), and 1-5 years (3), respectively. The unit of analysis is the child. Adjusted p-values (FT) refer to p-values of the coefficient on Full Treatment which are adjusted using the Benjamini-Hochberg method. *** p<0.01; ** p<0.05; * p<0.10

[66] An important assumption of the Cox model is that the relative effect of a covariate on the hazard function is constant over time (proportional hazard rates). We therefore do not include district fixed effects in our Cox models, since different regions of Uganda experience different seasonal patterns and thus different temporal patterns of child mortality rates. Our results are not affected by the exclusion of district fixed effects.

Table E25: Main outcomes – Principal component indices

	(1) **Utilization**	(2) **Treatment quality**	(3) **Patient satisfaction**	(4) **Health outcomes**
Full treatment	0.029	0.057	0.094***	-0.011
	(0.031)	(0.037)	(0.033)	(0.031)
Constant	-0.053**	-0.030	-0.047*	0.011
	(0.022)	(0.029)	(0.025)	(0.024)
N	7,288	7,288	7,288	4,212
R^2	0.284	0.040	0.028	0.042

Notes. Main outcome indices constructed using principal component analysis instead of averaged z-scores. Estimates from Equation 1 comparing outcomes measured at endline between the Full treatment arm and the Control. All models include district fixed effects and demeaned baseline covariates, as well as their interaction with the treatment indicator. Robust standard errors are clustered at the health center level. *** p<0.01; ** p<0.05; * p<0.10

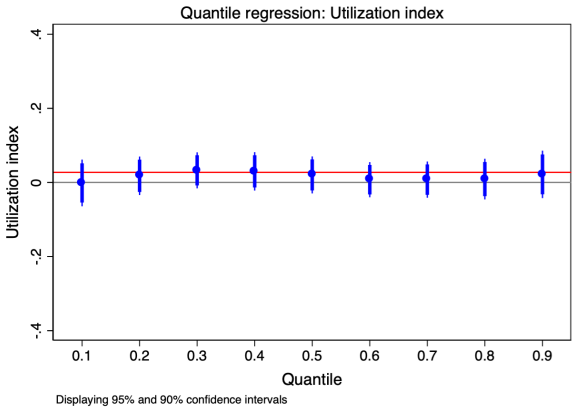
Table E26: Intermediate outcomes – Principal component indices

	(1) **Citizen knowledge**	(2) **HC staff knowledge**	(3) **Efficacy**	(4) **Community responsibility**	(5) **Community monitoring**	(6) **Relationship**	(7) **HC transparency**
Full treatment	-0.106***	0.175	-0.033	-0.017	-0.026	0.079**	0.018
	(0.032)	(0.140)	(0.036)	(0.027)	(0.032)	(0.035)	(0.115)
Constant	0.052**	-0.087	0.016	0.009	0.013	-0.041	0.005
	(0.023)	(0.096)	(0.024)	(0.019)	(0.022)	(0.025)	(0.079)
N	7,288	187	7,288	7,288	7,288	7,288	187
R^2	0.132	0.247	0.044	0.053	0.042	0.059	0.493

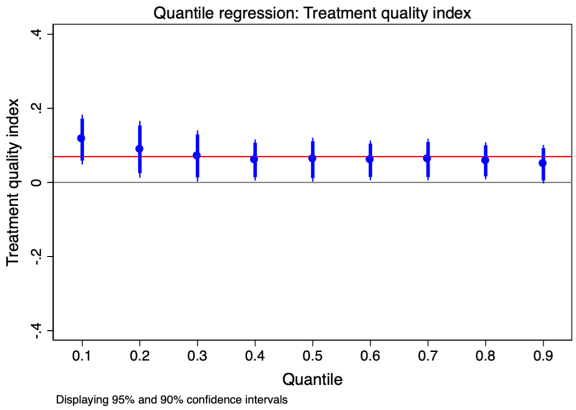
Notes. Intermediate outcome indices constructed using principal component analysis instead of averaged z-scores. Estimates from Equation 1 comparing outcomes measured at endline between the Full treatment arm and the Control. All models include district fixed effects and demeaned baseline covariates, as well as their interaction with the treatment indicator. Robust standard errors are clustered at the health center level. *** p<0.01; ** p<0.05; * p<0.10

Figure 6: Quantile regressions of treatment effects

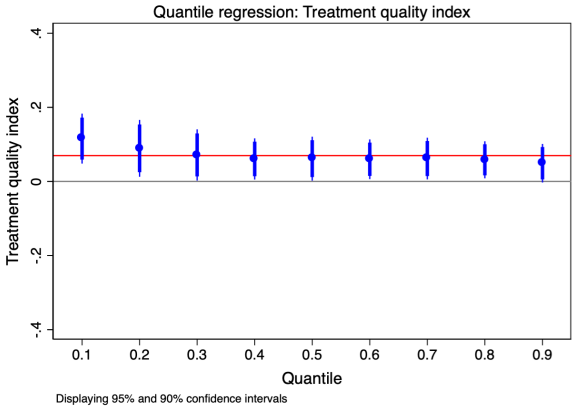
(a) Utilization



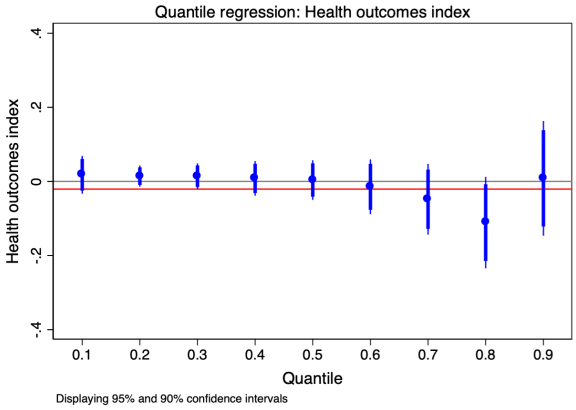
(b) Treatment quality



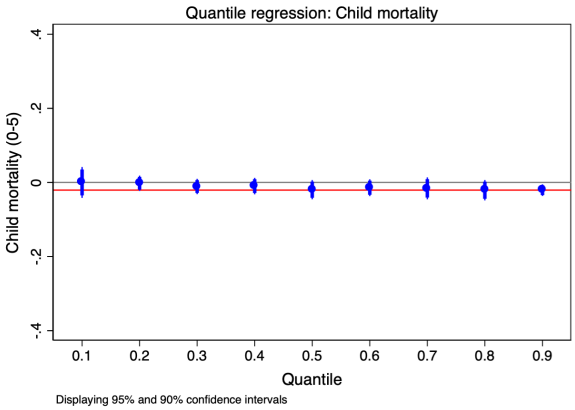
(c) Patient satisfaction



(d) Health outcomes



(e) Child mortality



Additional Robustness Checks on Treatment Quality

Table E27: Robustness check – Excluding subcomponents of treatment quality

	(1) Treatment quality wo/ used equipment	(2) Treatment quality wo/ waiting time	(3) Treatment quality wo/ examined by trained staff	(4) Treatment quality wo/ privacy during exam	(5) Treatment quality wo/ received test when needed	(6) Treatment quality wo/ diagnosis explained clearly	(7) Treatment quality wo/ % staff presence	(8) Treatment quality wo/ facility cleanliness	(9) Treatment quality wo/ drug availability
Full treatment	0.070***	0.021	0.142	0.000	0.015*	0.013	0.023*	0.017	0.016
	(0.026)	(0.013)	(2.829)	(0.002)	(0.008)	(0.014)	(0.013)	(0.041)	(0.019)
Constant	0.000	0.759***	69.736***	0.831***	0.910***	0.742***	0.697***	0.379***	0.691***
	(0.020)	(0.014)	(2.228)	(0.152)	(0.012)	(0.015)	(0.013)	(0.043)	(0.056)
N	7,288	7,288	7,288	7,288	7,288	7,288	7,288	187	187
R^2	0.102	0.023	0.084	0.026	0.039	0.030	0.019	0.299	0.166

Notes. Estimates from Equation 1 comparing outcomes between the Full treatment arm and the Control. All models include district fixed effects and demeaned baseline covariates, as well as their interaction with the treatment indicator. Robust standard errors are clustered at the health center level. The dependent variable is the treatment quality index, excluding one subcomponent at a time. *** p<0.01; ** p<0.05; * p<0.10

Table E28: Drugs availability in details

	(1) Health workers selling drugs on the side	(2) Patient did not receive drugs because no drug supplies
Full treatment	0.006	-0.018**
	(0.009)	(0.009)
Constant	0.883***	0.088***
	(0.006)	(0.007)
N	7,288	6,184
R^2	0.073	0.021

Notes. Estimates from Equation 1 comparing outcomes between the Full treatment arm and the Control. All models include district fixed effects and demeaned baseline covariates, as well as their interaction with the treatment indicator. Robust standard errors are clustered at the health center level. *** p<0.01; ** p<0.05; * p<0.10

Table E29: Robustness check main outcomes with treatment quality without HC-level subcomponents

	(1) **Utilization**	(2) **Treatment quality**	(3) **Patient satisfaction**	(4) **Health outcomes**	(5) **Child mortality**
A: Without control variables					
Program impact	0.034	0.038*	0.075***	-0.014	-0.010
	(0.023)	(0.020)	(0.025)	(0.028)	(0.008)
Constant	-0.022	-0.002	-0.001	-0.000	0.061***
	(0.017)	(0.015)	(0.018)	(0.022)	0.006
B: Without district fixed effects					
Program impact	0.013	0.036*	0.075***	-0.014	-0.009
	(0.028)	(0.020)	(0.025)	(0.029)	(0.008)
Constant	-0.012	-0.002	-0.001	0.000	0.063***
	(0.020)	(0.016)	(0.018)	(0.023)	0.006
C: With outcome measures aggregated at HC level					
Program impact	0.035*	0.031	0.080***	-0.012	-0.011
	(0.021)	(0.019)	(0.025)	(0.029)	(0.008)
Constant	-0.014	-0.002	-0.001	-0.009	0.061***
	(0.014)	(0.015)	(0.019)	(0.022)	(0.006)
D: Difference between post and pre-treatment values					
Program impact	0.004	0.039*	0.067***	-0.023	-0.008
	(0.025)	(0.022)	(0.025)	(0.034)	(0.008)
Constant	-0.023*	0.010	-0.011	0.028	0.045***
	(0.012)	(0.012)	(0.015)	(0.040)	(0.005)
Observations (A & B)	7,288	7,288	7,288	4,930	187
Observations (C)	187	187	187	187	187
Observations (D)	14,576	14,576	14,576	9,860	374

Notes. Estimates from Equation 1 comparing outcomes between the Full treatment arm and the Control, with the following variations: Panel A shows results without covariates, panel B without district fixed effects, and panel C aggregates outcome measures and covariates at the health center level (the unit of randomization). Panel D shows results from a difference in difference estimation. Here, the treatment quality index excludes the components measured at the health center level. *** p<0.01; ** p<0.05; * p<0.10

E.5 Additional Analyses

Mapping of Increases in Treatment Quality and Patient Satisfaction

Table E30: Treatment quality at midline and patient satisfaction at endline

	Difference treatment quality ML-BL $<= 0$	Difference treatment quality ML-BL > 0	P-value difference
Difference EL-BL Patient satisfaction	0.022	0.069	0.178

Notes. Sample restricted to the full treatment arm, the unit of observation is the health center catchment area (n=92). Two-sided t-test of the mean change in patient satisfaction between baseline and endline, by whether the change in treatment quality between baseline and midline was positive.

Presence of Local Government Officials

Table E31: Main outcomes and presence of an official

	(1) **Utilization**	(2) **Treatment quality**	(3) **Patient satisfaction**	(4) **Health outcomes**	(5) **Child mortality**
Full treatment	0.049**	0.054*	0.074***	-0.006	-0.011
	(0.023)	(0.028)	(0.026)	(0.032)	(0.009)
Subcounty official present	-0.068*	0.048	0.008	-0.047	0.001
	(0.036)	(0.036)	(0.038)	(0.040)	(0.013)
Constant	-0.020	0.000	-0.002	0.004	0.061***
	(0.016)	(0.020)	(0.018)	(0.021)	(0.006)
N	7,288	7,288	7,288	4,212	187
R^2	0.231	0.104	0.043	0.026	0.197
Full treatment + presence	-0.019	0.102***	0.082**	-0.053	-0.010
	(0.034)	(0.035)	(0.036)	(0.036)	(0.012)

Notes. Subcounty official present indicates whether an official from the local government was present at either the community dialogue or the interface meeting. Estimates from Equation 1 comparing outcomes between the Full treatment arm and the Control. The unit of observation in columns (1)-(4) is the household, in column (5) it is health center catchment area. All models include district fixed effects and demeaned baseline covariates, as well as their interaction with the treatment indicator. Robust standard errors are clustered at health center level. *** p<0.01; ** p<0.05; * p<0.10

Table E32: Subcomponents of the treatment quality index and presence of an official

	(1) **Treatment quality index**	(2) Used equipment	(3) Waiting time	(4) Examined by trained staff	(5) Privacy during exam	(6) Received test when needed	(7) Diagnosis explained clearly	(8) % staff presence	(9) Facility cleanliness	(10) Drug availability
Full treatment	0.054*	0.045	0.005	0.008	0.069	0.023	0.079**	-0.046	0.139	0.168
	(0.028)	(0.040)	(0.052)	(0.039)	(0.042)	(0.044)	(0.036)	(0.145)	(0.166)	(0.121)
Subcounty official present	0.048	0.033	-0.007	0.002	0.013	0.038	-0.066	0.321	-0.065	0.188
	(0.036)	(0.047)	(0.085)	(0.030)	(0.049)	(0.053)	(0.046)	(0.196)	(0.168)	(0.163)
Constant	0.000	0.031	-0.030	-0.000	-0.000	0.011	0.005	0.005	0.013	0.012
	(0.020)	(0.029)	(0.033)	(0.026)	(0.035)	(0.032)	(0.025)	(0.107)	(0.112)	(0.093)
N	7,288	7,288	7,288	7,288	7,288	7,288	7,288	187	187	187
R^2	0.104	0.023	0.084	0.026	0.039	0.030	0.020	0.310	0.167	0.425
Full treatment + presence	0.102***	0.079*	-0.003	0.010	0.081	0.061	0.013	0.276	0.073	0.357**
	(0.035)	(0.045)	(0.080)	(0.035)	(0.053)	(0.052)	(0.044)	(0.190)	(0.150)	(0.159)

Notes. Subcounty official present indicates whether an official from the local government was present at either the community dialogue or the interface meeting. All models include district fixed effects as well as demeaned baseline covariates and their interaction with the treatment indicator. Robust standard errors are clustered at the health center level. The dependent variable in column (1) is the averaged z-score index of the components presented in columns (2)-(10). The latter are z-scores of (2) whether household members reported that, during their most recent visit to the health center, equipment was used during examination, (3) waiting time consisting of the total amount of time spent by the household members waiting for the initial consultation and the examination; whether household members declared that, during their most recent visit to the health center, (4) they were examined by trained health care staff, (5) they had privacy during their examination, (6) lab tests were administered, (7) their diagnosis was clearly explained to them; (8) percent of staff in attendance during an unannounced visit to the health center, measured at the health center level, (9) condition of the clinic (cleanliness of floors and walls, whether the clinic smelled as observed during unannounced visit to health center), measured at the health center level, (10) share of months in which stock cards indicated availability of six key tracer drugs in the past three months, measured at the health center level. *** p<0.01; ** p<0.05; * p<0.10.

Table E33: Subcomponents of patient satisfaction and presence of an official

	(1) **Patient satisfaction index**	(2) Satisfied by HC quality	(3) Satisfied with quality of care	(4) Polite staff	(5) Staff interested in health	(6) Free to express clearly	(7) Availability of staff improving
Full treatment	0.074***	0.060*	0.075*	0.111***	0.077**	0.043	0.039
	(0.026)	(0.036)	(0.039)	(0.030)	(0.035)	(0.043)	(0.043)
Subcounty official present	0.008	0.005	-0.002	-0.031	-0.017	-0.008	0.119**
	(0.038)	(0.052)	(0.052)	(0.050)	(0.045)	(0.060)	(0.053)
Control	-0.002	0.017	0.016	0.020	0.013	0.020	-0.024
	(0.018)	(0.023)	(0.027)	(0.024)	(0.025)	(0.029)	(0.032)
N	7,288	7,288	7,288	7,288	7,288	7,288	7,288
R^2	0.043	0.044	0.024	0.018	0.023	0.019	0.055
Full treatment + presence	0.082**	0.065	0.073	0.080	0.060	0.035	0.158***
	(0.036)	(0.048)	(0.049)	(0.051)	(0.042)	(0.058)	(0.054)

Notes. Subcounty official present indicates whether an official from the local government was present at either the community dialogue or the interface meeting. All models include district fixed effects and demeaned baseline covariates, as well as their interaction with the treatment indicator. Robust standard errors are clustered at the health center level. The dependent variable in column (1) is the averaged z-score index of the outcomes presented in columns (2)-(8). The latter are z-scores of indicator variables of whether household members declared that (2) the services currently offered at the health center are of "very high quality" or "somewhat high quality", (3) they were "very satisfied" or "satisfied" with the quality of care received during their most recent visits to the health center, (4) during their most recent visit to the health center, the person conducting the examination behaved politely/showed respect, (5) during their most recent visit to the health center, the person conducting the examination appeared to be interested in their health condition, (6) during their most recent visit to the health center, the person conducting the examination listened to what they had to say, (7) during their most recent visit to the health center, they felt free to express themselves to the person conducting the examination, (8) compared to the year before, the availability of medical staff had improved at the health center. *** $p<0.01$; ** $p<0.05$; * $p<0.10$.

Alternative Utilization Index Including Deliveries and ANC Visits

Table E34: Utilization index including deliveries and ANC visits – Subcomponents

	(1) **Utilization index**	(2) Vaccination rates, children<36 months	(3) % of visits to HC, vs. other providers	(4) Number of visits to HC	(5) Delivering in the assigned HC	(6) Number of visits to ANC
Full treatment	0.028	0.053	0.034	-0.001	0.046	0.028
	(0.021)	(0.035)	(0.032)	(0.027)	(0.052)	(0.034)
Constant	-0.153***	-0.007	-0.001	-0.017	-0.249***	0.137**
	(0.016)	(0.025)	(0.023)	(0.020)	(0.056)	(0.065)
N	7,288	4,212	7,288	7,288	2,571	3,487
R^2	0.202	0.057	0.178	0.284	0.319	0.013
Adjusted p-value (FT)		0.517	0.517	0.981	0.517	0.517

Notes. Alternative operationalization of the utilization index including deliveries and ANC visits. Estimates are derived from from Equation 1, comparing outcomes between the Full treatment arm and the Control. All models include district fixed effects as well as demeaned baseline covariates and their interaction with the treatment indicator. Robust standard errors are clustered at the health center level. The dependent variable in column (1) is the averaged z-score index of the outcomes presented in columns (2)-(6). The latter are z-scores of (2) vaccination rates of children under 36 months, (3) share of visits to the designated health center versus other providers in the past 12 months, (4) number of visits to the designated health center in the past 12 months, (5) delivering in the designated health center, conditional on having had a delivery in the household in the past 24 months, and (6) the number of ANC visits, conditional on having had a pregnant household member in the past 24 months. Adjusted p-values (FT) refer to p-values of the coefficient on Full Treatment which are adjusted using the Benjamini-Hochberg method. All models include district fixed effects and demeaned baseline covariates, as well as their interaction with the treatment indicator. Robust standard errors are clustered at the health center level. *** p<0.01; ** p<0.05; * p<0.10

Table E35: Utilization index including deliveries and ANC visits (HC3s only) – Subcomponents

	(1) **Utilization index**	(2) Vaccination rates, children<36 months	(3) % of visits to HC, vs. other providers	(4) Number of visits to HC	(5) Delivering in the assigned HC	(6) Number of visits to ANC
Full treatment	0.040	-0.010	0.035	0.025	0.074	0.053
	(0.035)	(0.049)	(0.051)	(0.036)	(0.080)	(0.053)
Constant	-0.156***	0.030	-0.008	-0.054**	-0.030	0.199*
	(0.025)	(0.029)	(0.038)	(0.025)	(0.101)	(0.109)
N	2,998	1,719	2,998	2,998	1,041	1,451
R^2	0.196	0.045	0.173	0.275	0.205	0.024
Adjusted p-value (FT)		0.845	0.625	0.625	0.625	0.625

Notes. Alternative operationalization of the utilization index including deliveries and ANC visits. The sample is restricted to HC3s. Estimates are derived from from Equation 1, comparing outcomes between the Full treatment arm and the Control. All models include district fixed effects as well as demeaned baseline covariates and their interaction with the treatment indicator. Robust standard errors are clustered at the health center level. The dependent variable in column (1) is the averaged z-score index of the outcomes presented in columns (2)-(6). The latter are z-scores of (2) vaccination rates of children under 36 months, (3) share of visits to the designated health center versus other providers in the past 12 months, (4) number of visits to the designated health center in the past 12 months, (5) delivering in the designated health center, conditional on having had a delivery in the household in the past 24 months, and (6) the number of ANC visits, conditional on having had a pregnant household member in the past 24 months. Adjusted p-values (FT) refer to p-values of the coefficient on Full Treatment which are adjusted using the Benjamini-Hochberg method. All models include district fixed effects and demeaned baseline covariates, as well as their interaction with the treatment indicator. Robust standard errors are clustered at the health center level. *** p<0.01; ** p<0.05; * p<0.10

E.6 Main Results, by Treatment Arm

Main Specification

Table E36: Utilization outcomes – All treatment arms

	(1) **Utilization index**	(2) Vaccination rates, children<36 months	(3) % of visits to HC, vs. other providers	(4) Number of visits to HC
Full treatment	0.027	0.048	0.038	-0.002
	(0.022)	(0.036)	(0.033)	(0.027)
Information and mobilization only	0.013	0.026	0.025	-0.003
	(0.022)	(0.034)	(0.031)	(0.029)
Interface only	0.054**	0.049	0.074**	0.038
	(0.022)	(0.034)	(0.033)	(0.027)
Constant	-0.018	-0.008	0.001	-0.016
	(0.015)	(0.025)	(0.022)	(0.020)
N	14,609	8,548	14,609	14,609
R^2	0.221	0.045	0.173	0.273
P-value (Info/mobilization = Interface)	0.066	0.485	0.151	0.141
P-value (Info/mobilization = Full treatment)	0.518	0.520	0.712	0.962
P-value (Interface = Full treatment)	0.234	0.979	0.296	0.121

Notes. Estimates comparing outcomes between each treatment arm and the control from Equation 2. Each treatment arm enters as an indicator variable. All models include district fixed effects and demeaned baseline covariates, as well as their interactions with the treatment indicators. Robust standard errors are clustered at the health center level. The dependent variable in column (1) is the averaged z-score index of columns (2)-(4). The remaining dependent variables are z-scores of: (2) vaccination rates of children under 36 months; (3) the share of visits to the designated health center versus other providers; (4) the number of visits to the designated health center by all household members. *** p<0.01; ** p<0.05; * p<0.10

Table E37: Treatment quality outcomes – All treatment arms

	(1) **Treatment quality index**	(2) Used equipment	(3) Waiting time	(4) Examined by trained staff	(5) Privacy during exam	(6) Received test when needed	(7) Diagnosis explained clearly	(8) % staff presence	(9) Facility cleanliness	(10) Drug availability
Full treatment	0.071***	0.062*	0.003	0.006	0.075*	0.041	0.063*	0.030	0.107	0.246**
	(0.026)	(0.037)	(0.049)	(0.032)	(0.040)	(0.041)	(0.033)	(0.136)	(0.137)	(0.109)
Information and mobilization only	0.013	-0.035	-0.011	0.004	0.050	-0.007	-0.019	-0.121	-0.074	0.313***
	(0.029)	(0.040)	(0.043)	(0.039)	(0.043)	(0.046)	(0.036)	(0.140)	(0.135)	(0.111)
Interface only	0.022	0.036	-0.035	0.017	0.015	0.036	0.022	-0.138	-0.099	0.343***
	(0.027)	(0.042)	(0.046)	(0.033)	(0.045)	(0.046)	(0.037)	(0.134)	(0.137)	(0.103)
Constant	-0.002	0.021	-0.035	-0.001	0.001	0.012	-0.003	0.013	0.016	0.010
	(0.021)	(0.028)	(0.034)	(0.026)	(0.035)	(0.032)	(0.025)	(0.106)	(0.109)	(0.088)
N	14,609	14,609	14,609	14,609	14,609	14,609	14,609	376	376	376
R^2	0.102	0.030	0.102	0.015	0.031	0.035	0.021	0.299	0.176	0.367
P-value (Info/mobilization = Interface)	0.740	0.098	0.558	0.677	0.382	0.369	0.293	0.892	0.835	0.748
P-value (Info/mobilization = Full treatment)	0.025	0.011	0.751	0.956	0.475	0.265	0.019	0.239	0.132	0.525
P-value (Interface = Full treatment)	0.032	0.521	0.417	0.616	0.106	0.895	0.247	0.163	0.096	0.292

Notes. Estimates comparing outcomes between each treatment arm and the control from Equation 2. Each treatment arm enters as an indicator variable. All models include district fixed effects and demeaned baseline covariates, as well as their interactions with the treatment indicators. Robust standard errors are clustered at the health center level. The dependent variable in column (1) is the averaged z-score index of columns (2)-(10). The latter are z-scores of (2) whether household members reported that, during their most recent visit to the health center, equipment was used during examination, (3) waiting time consisting of the total amount of time spent by the household members waiting for the initial consultation and the examination; whether household members declared that, during their most recent visit to the health center, (4) they were examined by trained health care staff, (5) they had privacy during their examination, (6) lab tests were administered, (7) their diagnosis was clearly explained to them; (8) percent of staff in attendance during an unannounced visit to the health center, measured at the health center level, (9) condition of the clinic (cleanliness of floors and walls, whether the clinic smelled as observed during unannounced visit to health center), measured at the health center level, (10) share of months in which stock cards indicated availability of six key tracer drugs in the past three months, measured at the health center level. *** p<0.01; ** p<0.05; * p<0.10

Table E38: Patient satisfaction outcomes – All treatment arms

	(1) **Patient satisfaction index**	(2) Satisfied by HC quality	(3) Satisfied with quality of care	(4) Polite staff	(5) Staff interested in health	(6) Staff listening	(7) Free to express clearly	(8) Availability of staff improving
Full treatment	0.080***	0.105***	0.063*	0.074**	0.104***	0.076**	0.043	0.088**
	(0.024)	(0.035)	(0.034)	(0.036)	(0.030)	(0.031)	(0.040)	(0.041)
Information and mobilization only	0.073***	0.100***	0.056	0.054	0.114***	0.102***	0.077*	0.018
	(0.026)	(0.036)	(0.035)	(0.041)	(0.034)	(0.033)	(0.040)	(0.043)
Interface only	0.064***	0.067**	0.060*	0.048	0.093***	0.062**	0.089**	0.025
	(0.022)	(0.034)	(0.033)	(0.033)	(0.028)	(0.030)	(0.037)	(0.044)
Constant	-0.006	-0.011	0.006	0.005	0.009	0.007	0.008	-0.027
	(0.018)	(0.023)	(0.023)	(0.027)	(0.023)	(0.025)	(0.030)	(0.032)
N	14,609	14,609	14,609	14,609	14,609	14,609	14,609	14,609
R^2	0.040	0.071	0.040	0.019	0.016	0.020	0.016	0.057
P-value (Info/mobilization = Interface)	0.697	0.356	0.896	0.870	0.464	0.182	0.728	0.878
P-value (Info/mobilization = Full treatment)	0.778	0.905	0.839	0.608	0.740	0.403	0.360	0.075
P-value (Interface = Full treatment)	0.441	0.285	0.934	0.397	0.662	0.612	0.174	0.109

Notes. Estimates comparing outcomes between each treatment arm and the control from Equation 2. Each treatment arm enters as an indicator variable. All models include district fixed effects and demeaned baseline covariates, as well as their interactions with the treatment indicators. Robust standard errors are clustered at the health center level. The dependent variable in column (1) is the averaged z-score index of columns (2)-(8). (The latter are z-scores of indicator variables of whether household members declared that (2) the services currently offered at the health center are of "very high quality" or "somewhat high quality", (3) they were "very satisfied" or "satisfied" with the quality of care received during their most recent visits to the health center, (4) during their most recent visit to the health center, the person conducting the examination behaved politely/showed respect, (5) during their most recent visit to the health center, the person conducting the examination appeared to be interested in their health condition, (6) during their most recent visit to the health center, the person conducting the examination listened to what they had to say, (7) during their most recent visit to the health center, they felt free to express themselves to the person conducting the examination, (8) compared to the year before, the availability of medical staff had improved at the health center. *** p<0.01; ** p<0.05; * p<0.10

Table E39: Health outcomes at the child level – All treatment arms

	(1) **Health outcomes index**	(2) **Weight/Age 0-18 months**	(3) **MUAC 0-18 months**	(4) **Weight/Age 18-36 months**	(5) **MUAC 18-36 months**
Full treatment	-0.003	-0.001	-0.013	0.003	0.019
	(0.028)	(0.048)	(0.048)	(0.032)	(0.028)
Information and mobilization only	-0.023	-0.036	-0.026	-0.011	-0.011
	(0.029)	(0.048)	(0.048)	(0.035)	(0.029)
Interface only	-0.011	-0.019	-0.023	0.014	0.026
	(0.028)	(0.046)	(0.046)	(0.036)	(0.032)
Constant	-0.488***	0.005	0.005	-0.461***	-0.630***
	(0.022)	(0.038)	(0.038)	(0.026)	(0.020)
N	10,023	4,379	4,379	5,644	5,644
R^2	0.103	0.011	0.012	0.207	0.328
P-value (Info/mobilization = Interface)	0.653	0.661	0.933	0.500	0.258
P-value (Info/mobilization = Full treatment)	0.472	0.408	0.753	0.687	0.303
P-value (Interface = Full treatment)	0.767	0.651	0.805	0.739	0.840

Notes. Estimates comparing outcomes between each treatment arm and the control from Equation 2. Each treatment arm enters as an indicator variable. All models include district fixed effects and demeaned baseline covariates, as well as their interactions with the treatment indicators. Robust standard errors are clustered at the health center level. The dependent variable in column (1) is the averaged z-score index of columns (2)-(7). The latter are z-scores of (2) the average ratio of weight over number of months for children under 18 months, (3) the average ratio of weight over number of months for children 18-36 months old, (4) the average ratio of upper arm circumference over number of months for children under 18 months, (5) the average ratio of upper arm circumference over number of months for children 18-36 months old. *** p<0.01; ** p<0.05; * p<0.10

Table E40: Child mortality at the HC level

	(1) Child mortality 0-5 years old	(2) Child mortality 0-1 year old	(3) Child mortality 1-5 years old
Full treatment	-0.011	-0.006	-0.006
	(0.008)	(0.007)	(0.004)
Information and mobilization only	-0.020**	-0.013**	-0.007
	(0.008)	(0.006)	(0.004)
Interface only	-0.009	-0.004	-0.006
	(0.008)	(0.007)	(0.005)
Constant	0.061***	0.042***	0.020***
	(0.006)	(0.005)	(0.003)
N	376	376	376
R^2	0.151	0.160	0.159
P-value (Info/mobilization = Interface)	0.204	0.212	0.753
P-value (Info/mobilization = Full treatment)	0.281	0.334	0.738
P-value (Interface = Full treatment)	0.820	0.797	0.992

Notes. Estimates comparing outcomes between each treatment arm and the control from Equation 2. Each treatment arm enters as an indicator variable. All models include district fixed effects and demeaned baseline covariates, as well as their interactions with the treatment indicators. The dependent variable is the child mortality rate in the health center catchment area calculated with the synthetic cohort approach, in the age bracket 0-5 years (1), 0-12 months (2), and 1-5 years (3), respectively. The unit of analysis is the health center catchment area. *** $p<0.01$; ** $p<0.05$; * $p<0.10$

Table E41: Intermediates outcomes – All treatments

	(1) **Citizen knowledge**	(2) **HC staff knowledge**	(3) **Efficacy**	(4) **Community responsibility**	(5) **Community monitoring**	(6) **Relationship**	(7) **HC transparency**
Full treatment	-0.054**	0.140	-0.019	-0.011	0.003	0.039	-0.026
	(0.024)	(0.118)	(0.023)	(0.021)	(0.029)	(0.040)	(0.077)
Information and mobilization only	0.001	0.226**	0.006	0.001	0.036	0.063	-0.126
	(0.024)	(0.110)	(0.023)	(0.022)	(0.030)	(0.039)	(0.077)
Interface only	-0.031	0.107	-0.011	-0.025	0.041	0.020	-0.104
	(0.023)	(0.110)	(0.022)	(0.021)	(0.029)	(0.042)	(0.080)
Constant	-0.019	-0.012	-0.003	-0.005	0.005	0.004	-0.010
	(0.017)	(0.079)	(0.015)	(0.015)	(0.020)	(0.029)	(0.055)
N	14,609	376	14,609	14,609	14,609	14,609	376
R^2	0.193	0.191	0.044	0.047	0.077	0.050	0.402
P-value (Info/mobilization = Interface)	0.179	0.266	0.467	0.240	0.874	0.296	0.784
P-value (Info/mobilization = Full treatment)	0.032	0.457	0.296	0.566	0.284	0.536	0.198
P-value (Interface = Full treatment)	0.327	0.772	0.708	0.510	0.200	0.652	0.336

Notes. Estimates comparing outcomes between each treatment arm and the control group for intermediate outcome indices from Equation 2. Each treatment arm enters as an indicator variable. All models include district fixed effects and demeaned baseline covariates, as well as their interactions with the treatment indicators. Robust standard errors are clustered at the health center level. *** p<0.01; ** p<0.05; * p<0.10

Pre-specified Specification

In addition, we show the main results from the pre-specified model, which includes an interaction term rather than estimating average treatment effects for each cell more flexibly, below.

Table E42: Main outcomes - All treatment arms: Averaged z-score indices

	(1) **Utilization**	(2) **Treatment quality**	(3) **Patient satisfaction**	(4) **Health outcomes**	(5) **Child mortality**
Information	0.013	0.013	0.073***	-0.023	-0.020**
	(0.022)	(0.029)	(0.026)	(0.029)	(0.008)
Interface	0.054**	0.022	0.064***	-0.011	-0.009
	(0.022)	(0.027)	(0.022)	(0.028)	(0.008)
Information x Interface	-0.039	0.036	-0.058*	0.030	0.018
	(0.031)	(0.037)	(0.033)	(0.039)	(0.011)
Constant	-0.018	-0.002	-0.006	-0.488***	0.061***
	(0.015)	(0.021)	(0.018)	(0.022)	(0.006)
N	14,609	14,609	14,609	10,023	376
R^2	0.221	0.102	0.040	0.103	0.151
Informantion + Information x Interface	-0.026	0.049**	0.016	0.008	-0.002
	(0.022)	(0.023)	(0.020)	(0.025)	(0.008)
P-value (Information = Interface)	0.066	0.740	0.697	0.653	0.204

Notes. Estimates comparing outcomes between each treatment arm and the control. All models include district fixed effects as well as demeaned baseline covariates and their interaction with the treatment indicator. Robust standard errors are clustered at the health center level. *** p<0.01; ** p<0.05; * p<0.10

Table E43: Intermediate outcome indices – All treatment arms

	(1) **Citizen knowledge**	(2) **HC staff knowledge**	(3) **Efficacy**	(4) **Community responsibility**	(5) **Community monitoring**	(6) **Relationship**	(7) **HC transparency**
Information	0.001	0.245**	0.005	0.004	0.039	0.066*	-0.111
	0.025	0.114	0.023	0.022	0.030	0.039	0.083
Interface	-0.028	0.086	-0.010	-0.023	0.042	0.020	-0.115
	0.023	0.113	0.022	0.021	0.029	0.043	0.084
Information x Interface	-0.027	-0.191	-0.016	0.008	-0.077*	-0.046	0.187
	0.034	0.167	0.032	0.030	0.042	0.057	0.120
Constant	-0.019	-0.011	-0.003	-0.005	0.004	0.004	-0.009
	0.017	0.079	0.015	0.015	0.020	0.029	0.054
N	14,609	376	14,609	14,609	14,609	14,609	376
R^2	0.194	0.195	0.045	0.049	0.079	0.051	0.411
Information + Information x Interface	-0.025	0.054	-0.011	0.012	-0.038	0.020	0.076
	(0.023)	(0.122)	(0.023)	(0.020)	(0.029)	(0.042)	(0.085)
P-value (Information = Interface)	0.210	0.167	0.514	0.230	0.912	0.273	0.966

Notes. Estimates comparing outcomes between each treatment arm and the control group for intermediate outcome indices. All models include district fixed effects and demeaned baseline covariates, as well as their interaction with the treatment indicator. Robust standard errors are clustered at the health center level. *** p<0.01; ** p<0.05; * p<0.10

E.7 Midline Results

The tables that follow show the treatment effect on main and intermediate outcome indices at midline, both for the comparison between treatment and the full intervention and for all treatment arms.

Main Results

Table E44: Main outcomes (midline)

	(1) **Utilization**	(2) **Treatment quality**	(3) **Patient satisfaction**	(4) **Health outcomes**	(5) **Child mortality**
Full treatment	-0.012	0.060*	0.036	0.024	-0.014
	(0.020)	(0.033)	(0.029)	(0.027)	(0.010)
Constant	-0.012	-0.002	-0.000	-0.153***	0.054***
	(0.015)	(0.024)	(0.019)	(0.023)	(0.008)
N	7,204	7,204	7,204	5,337	187
R^2	0.255	0.095	0.049	0.102	0.204
P-value (Full treatment = 0)	0.560	0.066	0.207	0.382	0.166
Adjusted p-value (FT)	0.560	0.331	0.344	0.477	0.344

Notes. Estimates from Equation 1 comparing outcomes measured at midline between the Full treatment arm and the Control. All models include district fixed effects and demeaned baseline covariates, as well as their interaction with the treatment indicator. Robust standard errors are clustered at the health center level. Adjusted p-values (FT) refer to p-values of the coefficient on Full Treatment which are adjusted using the Benjamini-Hochberg method. *** p<0.01; ** p<0.05; * p<0.10

Table E45: Treatment quality index (midline) and its subcomponents

	(1) **Treatment quality index**	(2) Used equipment	(3) Waiting time	(4) Examined by trained staff	(5) Privacy during exam	(6) Received test when needed	(7) Diagnosis explained clearly	(8) % staff presence	(9) Facility cleanliness	(10) Drug availability
Full treatment	0.060*	0.095*	0.076	-0.030	0.062	0.061	-0.046	0.075	-0.039	0.303***
	(0.033)	(0.051)	(0.063)	(0.044)	(0.038)	(0.054)	(0.041)	(0.142)	(0.152)	(0.107)
Constant	-0.002	0.029	-0.026	0.004	0.007	0.025	0.008	-0.055	0.010	-0.010
	(0.024)	(0.039)	(0.042)	(0.023)	(0.031)	(0.043)	(0.029)	(0.112)	(0.103)	(0.087)
N	7,204	7,204	7,204	7,204	7,204	7,204	7,204	187	187	187
R^2	0.095	0.034	0.150	0.012	0.026	0.064	0.022	0.212	0.235	0.476
Adjusted p-value (FT)	0.331	0.278	0.397	0.644	0.325	0.397	0.397	0.674	0.798	0.045

Notes. Estimates from Equation 1 comparing outcomes measured at midline between the Full treatment arm and the Control. All models include district fixed effects and demeaned baseline covariates, as well as their interaction with the treatment indicator. Robust standard errors are clustered at the health center level. Adjusted p-values (FT) refer to p-values of the coefficient on Full Treatment which are adjusted using the Benjamini-Hochberg method. *** p<0.01; ** p<0.05; * p<0.10

Table E46: Intermediate outcomes (midline) – Averaged z-score indices

	(1) **Citizen knowledge**	(2) **HC staff knowledge**	(3) **Efficacy**	(4) **Community responsibility**	(5) **Community monitoring**	(6) **Relationship**	(7) **HC transparency**
Full treatment	0.009	0.084	0.039**	0.009	0.018	-0.041	-0.090
	(0.023)	(0.099)	(0.019)	(0.021)	(0.024)	(0.043)	(0.064)
Constant	-0.009	-0.006	-0.001	-0.005	-0.001	0.003	-0.016
	(0.017)	(0.072)	(0.014)	(0.015)	(0.018)	(0.026)	(0.050)
N	7,204	187	7,204	7,204	7,204	7,204	187
R^2	0.157	0.442	0.044	0.044	0.050	0.095	0.352
P-value (Full treatment = 0)	0.693	0.395	0.045	0.684	0.470	0.339	0.163
Adjusted p-value (FT)	0.693	0.658	0.318	0.693	0.658	0.658	0.570

Notes. Estimates from Equation 1 comparing outcomes between the Full treatment arm and the Control for intermediate outcome indices measured at midline. All models include district fixed effects and demeaned baseline covariates, as well as their interaction with the treatment indicator. Robust standard errors are clustered at the health center level. Adjusted p-values (FT) refer to p-values of the coefficient on Full Treatment which are adjusted using the Benjamini-Hochberg method. *** p<0.01; ** p<0.05; * p<0.10

Table E47: Robustness check main outcomes (midline)

	(1) **Utilization**	(2) **Treatment quality**	(3) **Patient satisfaction**	(4) **Health outcomes**	(5) **Child mortality**
A: Without control variables					
Program impact	-0.006	0.057*	0.035	0.021	-0.013
	(0.022)	(0.034)	(0.029)	(0.028)	(0.010)
Constant	-0.015	-0.001	0.001	0.000	0.055***
	(0.017)	(0.025)	(0.020)	(0.020)	0.008
B: Without district fixed effects					
Program impact	-0.018	0.062*	0.038	0.021	-0.012
	(0.026)	(0.035)	(0.030)	(0.029)	(0.010)
Constant	-0.009	-0.003	-0.001	-0.000	0.055***
	(0.018)	(0.025)	(0.021)	(0.020)	0.008
C: With outcome measures aggregated at HC level					
Program impact	-0.011	0.050	0.039	0.024	-0.014
	(0.021)	(0.035)	(0.030)	(0.028)	(0.010)
Constant	-0.003	0.002	0.000	0.003	0.054***
	(0.015)	(0.025)	(0.021)	(0.020)	(0.008)
D: Difference between post and pre-treatment values					
Program impact	-0.035	0.065*	0.029	0.020	-0.011
	(0.024)	(0.035)	(0.030)	(0.028)	(0.010)
Constant	-0.023**	0.021	-0.008	0.014	0.045***
	(0.012)	(0.017)	(0.015)	(0.020)	(0.005)
Observations (A & B)	7,204	7,204	7,204	5,337	187
Observations (C)	187	187	187	187	187
Observations (D)	14,408	14,408	14,408	10,674	374

Notes. Estimates from Equation 1 comparing outcomes between the Full treatment arm and the Control, with the following variations: Panel A shows results without covariates, panel B without district fixed effects, and panel C aggregates outcome measures and covariates at the health center level (the unit of randomization). Panel D shows results from a difference in difference estimation. *** p<0.01; ** p<0.05; * p<0.10

Table E48: Robustness check treatment quality subcomponents by subcomponents (midline)

	(1) Treatment quality wo/ used equipment	(2) Treatment quality wo/ waiting time	(3) Treatment quality wo/ examined by trained staff	(4) Treatment quality wo/ privacy during exam	(5) Treatment quality wo/ received test when needed	(6) Treatment quality wo/ diagnosis explained clearly	(7) Treatment quality wo/ % staff presence	(8) Treatment quality wo/ facility cleanliness	(9) Treatment quality wo/ drug availability
Full treatment	0.057	0.057	0.071**	0.061*	0.062*	0.075**	0.059*	0.072**	0.032
	(0.035)	(0.036)	(0.036)	(0.036)	(0.034)	(0.035)	(0.031)	(0.028)	(0.033)
Constant	-0.003	-0.000	-0.002	-0.003	-0.002	-0.002	-0.000	-0.000	-0.006
	(0.025)	(0.026)	(0.027)	(0.026)	(0.025)	(0.026)	(0.023)	(0.021)	(0.022)
N	7,204	7,204	7,204	7,204	7,204	7,204	7,204	7,204	7,204
R^2	0.109	0.072	0.114	0.094	0.130	0.123	0.078	0.072	0.087

Table E49: Robustness check main outcomes with treatment quality without HC-level subcomponents (midline)

	(1) **Utilization**	(2) **Treatment quality**	(3) **Patient satisfaction**	(4) **Health outcomes**	(5) **Child mortality**
A: Without control variables					
Program impact	-0.006	0.035	0.035	0.021	-0.013
	(0.022)	(0.026)	(0.029)	(0.028)	(0.010)
Constant	-0.015	-0.003	0.001	0.000	0.055***
	(0.017)	(0.018)	(0.020)	(0.020)	0.008
B: Without district fixed effects					
Program impact	-0.018	0.030	0.038	0.021	-0.012
	(0.026)	(0.026)	(0.030)	(0.029)	(0.010)
Constant	-0.009	-0.001	-0.001	-0.000	0.055***
	(0.018)	(0.019)	(0.021)	(0.020)	0.008
C: With outcome measures aggregated at HC level					
Program impact	-0.011	0.022	0.039	0.024	-0.014
	(0.021)	(0.025)	(0.030)	(0.028)	(0.010)
Constant	-0.003	0.001	0.000	0.003	0.054***
	(0.015)	(0.017)	(0.021)	(0.020)	(0.008)
D: Difference between post and pre-treatment values					
Program impact	-0.035	0.035	0.029	0.020	-0.011
	(0.024)	(0.027)	(0.030)	(0.028)	(0.010)
Constant	-0.023**	0.010	-0.008	0.014	0.045***
	(0.012)	(0.012)	(0.015)	(0.020)	(0.005)
Observations (A & B)	7,204	7,204	7,204	5,337	187
Observations (C)	187	187	187	187	187
Observations (D)	14,408	14,408	14,408	10,674	374

Notes. Estimates from Equation 1 comparing outcomes between the Full treatment arm and the Control, with the following variations: Panel A shows results without covariates, panel B without district fixed effects, and panel C aggregates outcome measures and covariates at the health center level (the unit of randomization). Panel D shows results from a difference in difference estimation. Here, the treatment quality index excludes subcomponents measured at the HC level. *** p<0.01; ** p<0.05; * p<0.10

Table E50: Robustness check intermediate outcomes (midline)

	(1) **Citizen knowledge**	(2) **HC staff knowledge**	(3) **Efficacy**	(4) **Community responsibility**	(5) **Community monitoring**	(6) **Relationship**	(7) **HC transparency**
A: Without control variables							
Program impact	-0.001	0.046	0.038*	0.007	0.018	-0.041	-0.117*
	(0.031)	(0.103)	(0.020)	(0.022)	(0.024)	(0.045)	(0.070)
Constant	-0.005	0.010	-0.000	-0.004	-0.001	0.002	-0.001
	(0.022)	(0.076)	(0.014)	(0.016)	(0.018)	(0.027)	(0.058)
B: Without district fixed effects							
Program impact	-0.007	0.088	0.037	-0.004	0.013	-0.038	-0.103
	(0.037)	(0.110)	(0.023)	(0.027)	(0.027)	(0.044)	(0.064)
Constant	-0.001	-0.009	0.000	0.001	0.001	0.001	-0.011
	(0.025)	(0.083)	(0.017)	(0.019)	(0.021)	(0.027)	(0.051)
C: Without ouctome measures aggregated at HC level							
Program impact	0.003	0.084	0.036*	0.006	0.009	-0.017	-0.090
	(0.027)	(0.099)	(0.022)	(0.024)	(0.028)	(0.050)	(0.064)
Constant	-0.004	-0.006	-0.000	-0.002	0.003	-0.006	-0.016
	(0.019)	(0.072)	(0.015)	(0.017)	(0.022)	(0.033)	(0.050)
D: Difference between post and pre-treatment values							
Program impact	-0.005	0.117	0.035*	0.002	0.021	-0.060	-0.109*
	(0.038)	(0.104)	(0.021)	(0.022)	(0.027)	(0.046)	(0.065)
Constant	-0.003	0.044	-0.016	-0.013	0.027	-0.042	0.001
	(0.019)	(0.057)	(0.011)	(0.010)	(0.020)	(0.026)	(0.041)
Observations (A & B)	7,204	187	7,204	7,204	7,204	7,204	187
Observations (C)	187	187	187	187	187	187	187
Observations (D)	14,408	374	14,408	14,408	14,408	14,408	374

Midline Results by Treatment Arm

Table E51: Main outcomes (midline) – All treatments

	(1) **Utilization**	(2) **Treatment quality**	(3) **Patient satisfaction**	(4) **Health outcomes**	(5) **Child mortality**
Full treatment	-0.015	0.061*	0.039	0.025	-0.014
	(0.021)	(0.033)	(0.029)	(0.028)	(0.010)
Information and mobilization only	-0.031	0.066**	0.044	-0.015	-0.021**
	(0.022)	(0.033)	(0.027)	(0.029)	(0.009)
Interface only	-0.007	0.031	0.018	0.010	-0.011
	(0.023)	(0.033)	(0.028)	(0.029)	(0.010)
Constant	-0.009	-0.005	-0.003	-0.147***	0.054***
	(0.016)	(0.024)	(0.020)	(0.022)	(0.007)
N	14,459	14,459	14,459	10,787	376
R^2	0.239	0.091	0.048	0.093	0.153
P-value (Info/mobilization = Interface)	0.275	0.263	0.342	0.408	0.269
P-value (Info/mobilization = Full treatment)	0.422	0.871	0.849	0.154	0.392
P-value (Interface = Full treatment)	0.708	0.344	0.470	0.586	0.812

Notes. Estimates comparing midline outcomes between each treatment arm and the control. Each treatment arm enters as a separate indicator. All models include district fixed effects and demeaned baseline covariates, as well as their interactions with the treatment indicators. Robust standard errors are clustered at the health center level. *** p<0.01; ** p<0.05; * p<0.10

E.8 Results from T-Tests

This section shows results from two-sided t-tests of difference of means, with the unit of observation being the health center catchment area.

Table E52: Twosided t-tests (midline and endline)

	(1) Full treatment	(2) Information only	(3) Interface only	(4) Control	(5) P-value difference (1) - (4)	(6) P-value difference (1) & (2) - (3) & (4)	(7) P-value difference (1) & (3) - (2) & (4)
A. Midline levels of main outcome indices							
Utilization	5.96	5.88	6.28	6.22	0.37	0.08	0.70
Treatment quality	0.77	0.76	0.76	0.75	0.18	0.14	0.67
Patient satisfaction	0.77	0.77	0.77	0.76	0.26	0.17	0.80
Health outcomes	1.03	1.01	1.05	0.98	0.22	0.97	0.06
Mortality	0.04	0.04	0.05	0.06	0.26	0.12	0.90
B. Midline levels of intermediate outcome indices							
Citizen knowledge	0.40	0.39	0.40	0.40	0.84	0.88	0.89
Health worker knowledge	0.40	0.38	0.37	0.38	0.47	0.45	0.84
Efficacy	0.64	0.64	0.65	0.63	0.16	0.72	0.13
Community responsibility	0.28	0.29	0.28	0.28	0.91	0.97	0.84
Community monitoring	0.77	0.78	0.77	0.76	0.33	0.42	0.51
Relationship between health workers and community	0.80	0.82	0.81	0.81	0.44	0.70	0.12
Health center transparency	0.19	0.22	0.23	0.22	0.22	0.27	0.58
C. Endline levels of main outcome indices							
Utilization	6.16	6.19	6.47	6.36	0.55	0.27	0.83
Treatment quality	0.79	0.77	0.78	0.77	0.01	0.27	0.03
Patient satisfaction	0.78	0.78	0.77	0.76	0.01	0.03	0.15
Health outcomes	1.07	1.05	1.06	1.08	0.86	0.80	0.97
Mortality	0.05	0.05	0.06	0.06	0.31	0.11	0.89
D. Endline levels of intermediate outcome indices							
Citizen knowledge	0.43	0.44	0.43	0.45	0.12	0.45	0.13
Health worker knowledge	0.35	0.36	0.34	0.32	0.29	0.12	0.95
Efficacy	0.65	0.66	0.65	0.66	0.29	0.66	0.29
Community responsibility	0.33	0.33	0.33	0.34	0.60	0.92	0.39
Community monitoring	0.73	0.74	0.74	0.73	0.72	0.91	0.52
Relationship between health workers and community	0.80	0.81	0.80	0.79	0.17	0.17	0.62
Health center transparency	0.28	0.24	0.24	0.30	0.68	0.72	0.78
N	92	92	97	95			

E.9 Multiple Comparison Corrections

Given the number of outcome variables in our study, multiple testing is a concern. Main tables also include corrected p-values for the average effect of the full treatment, calculated using the Benjamini and Hochberg (1995) False Discovery Rate correction, in the bottom panel. This simple step-up procedure is slightly less punitive than a Bonferroni correction since it focuses exclusively on correcting for the false discovery rate (type I errors).

For outcome indices, the family is defined as the set of main outcome indices or the set of intermediate outcome indices, respectively. For components of an index, the family is defined as the set of components of a given index.

E.10 Outcomes in Sub-Samples Matching P2P

Table E53: Average treatment effects across sub-samples

	(1) Full Sample		(2) HCs whose baseline CMR are w/in 1 SD of P2P's baseline levels		(3) HC3s only		(4) Intersection of (2) and (3)	
	Coeff.	Obs.	Coeff.	Obs.	Coeff.	Obs.	Coeff.	Obs.
Utilization index	0.027 (0.022)	7,288	0.030 (0.032)	2,860	0.030 (0.034)	2,998	0.004 (0.057)	1,129
Treatment quality index	0.070*** (0.026)	7,288	0.020 (0.040)	2,860	0.030 (0.039)	2,998	0.006 (0.054)	1,129
Patient satisfaction index	0.077*** (0.024)	7,288	0.104*** (0.038)	2,860	0.097*** (0.034)	2,998	0.097* (0.052)	1,129
Health outcomes index	-0.003 (0.027)	4,930	-0.047 (0.044)	1,953	0.025 (0.041)	2,014	-0.049 (0.064)	805
Child mortality (child level)	1.059 (0.239)	10,118	0.210*** (0.086)	4,028	1.056 (0.413)	4,077	0.205 (0.215)	1,588
Child mortality (HC level)	-0.011 (0.008)	187	-0.023** (0.010)	73	-0.016 (0.013)	77	-0.015 (0.023)	29
Efficacy index	-0.022 (0.023)	7,288	0.021 (0.036)	2,860	-0.033 (0.033)	2,998	-0.007 (0.054)	1,129
Community responsibility index	-0.012 (0.020)	7,288	0.023 (0.031)	2,860	-0.028 (0.029)	2,998	0.050 (0.049)	1,129
Community monitoring index	0.006 (0.028)	7,288	0.111** (0.050)	2,860	0.029 (0.038)	2,998	0.182** (0.068)	1,129

Notes. This summary table shows the coefficients on the treatment indicator for the five main outcome indices as well as the three intermediate outcome indices relating to community monitoring in four increasingly restrictive samples. Column (1) shows results for the full sample, column (2) for health centers with a baseline under five mortality rate within one standard deviation of that in Björkman and Svensson (2009), column (3) for HC3s only, and column (4) for the intersection between columns (2) and (3). All samples are restricted to the full treatment and control group. Unless noted otherwise, the unit of observation is the household. For *child mortality (child level)*, we display hazard ratios estimated with a Cox proportional hazards model. A hazard ratio below (above) 1 implies that the treatment led to lower (higher) mortality rates. All models include district fixed effects as well as demeaned baseline covariates and their interaction with the treatment indicator. Models in columns (3) and (4) exclude the three health center level covariates due to limited variation within HC3s. Robust standard errors are clustered at health center level. *** p<0.01; ** p<0.05; * p<0.10

Table E54: Robustness checks for HCs within 1 standard deviation of U5MR in P2P at baseline – Main outcomes

	(1) **Utilization**	(2) **Treatment quality**	(3) **Patient satisfaction**	(4) **Health outcomes**	(5) **Child mortality**
A: Without control variables					
Program impact	0.029	0.019	0.104***	-0.029	-0.026**
	(0.033)	(0.040)	(0.037)	(0.043)	(0.011)
Constant	-0.023	0.035	-0.015	0.003	0.033**
	(0.027)	(0.032)	(0.026)	(0.035)	(0.014)
B: Without district fixed effects					
Program impact	0.048	0.043	0.108***	-0.013	-0.024**
	(0.045)	(0.049)	(0.038)	(0.047)	(0.011)
Constant	-0.032	0.020	-0.016	-0.007	0.041***
	(0.034)	(0.039)	(0.026)	(0.036)	(0.015)
C: With outcome measures aggregated at HC level					
Program impact	0.044	0.024	0.098**	-0.031	-0.023**
	(0.035)	(0.047)	(0.041)	(0.044)	(0.010)
Constant	-0.009	0.028	-0.009	0.000	0.023
	(0.026)	(0.039)	(0.027)	(0.033)	(0.015)
D: Difference between post and pre-treatment values					
Program impact	0.007	0.038	0.090**	0.028	-0.020*
	(0.037)	(0.044)	(0.038)	(0.060)	(0.012)
Constant	-0.036*	0.043	-0.018	-0.017	0.115***
	(0.018)	(0.028)	(0.024)	(0.056)	(0.007)
Observations (A & B)	2,860	2,860	2,860	1,953	73
Observations (C)	73	73	73	73	73
Observations (D)	5,720	5,720	5,720	3,906	146

Notes. The sample is restricted to HCs with a baseline under five mortality rate within one standard deviation of that in Björkman and Svensson (2009). Estimates from Equation 1 comparing outcomes between the Full treatment arm and the Control, with the following variations: Panel A shows results without covariates, panel B without district fixed effects, and panel C aggregates outcome measures and covariates at the health center level (the unit of randomization). Panel D shows results from a difference in difference estimation. *** p<0.01; ** p<0.05; * p<0.10

Table E55: Robustness checks for HC3s within 1 standard deviation of U5MR in P2P at baseline – Main outcomes

	(1) **Utilization**	(2) **Treatment quality**	(3) **Patient satisfaction**	(4) **Health outcomes**	(5) **Child mortality**
A: Without control variables					
Program impact	0.003	0.005	0.101*	-0.051	-0.023
	(0.062)	(0.054)	(0.055)	(0.067)	(0.020)
Constant	-0.014	0.038	0.026	0.034	0.019
	(0.043)	(0.033)	(0.030)	(0.053)	(0.020)
B: Without district fixed effects					
Program impact	0.036	-0.011	0.093*	-0.027	-0.023
	(0.065)	(0.076)	(0.053)	(0.088)	(0.016)
Constant	-0.030	0.022	0.034	0.020	0.033
	(0.040)	(0.057)	(0.035)	(0.066)	(0.022)
C: With outcome measures aggregated at HC level					
Program impact	-0.092	-0.104	0.099	-0.080	-0.015
	(0.084)	(0.104)	(0.077)	(0.136)	(0.023)
Constant	0.058	0.033	0.047	0.046	0.021
	(0.046)	(0.051)	(0.030)	(0.093)	(0.042)
D: Difference between post and pre-treatment values					
Program impact	0.020	0.033	0.085	0.029	-0.028
	(0.066)	(0.054)	(0.053)	(0.106)	(0.021)
Constant	-0.091***	0.147***	-0.062*	-0.061	0.119***
	(0.026)	(0.033)	(0.032)	(0.067)	(0.012)
Observations (A & B)	1,129	1,129	1,129	805	29
Observations (C)	29	29	29	29	29
Observations (D)	2,258	2,258	2,258	1,610	58

Notes. The sample is restricted to HC3s with a baseline under five mortality rate within one standard deviation of that in Björkman and Svensson (2009). Estimates from Equation 1 comparing outcomes between the Full treatment arm and the Control, with the following variations: Panel A shows results without covariates, panel B without district fixed effects, and panel C aggregates outcome measures and covariates at the health center level (the unit of randomization). Panel D shows results from a difference in difference estimation. *** p<0.01; ** p<0.05; * p<0.10

Table E56: Robustness checks with restricted samples – Community monitoring

	(1) **Full sample** **Full treatment & control**	(2) **Under 1 SD U5MR** **P2P baseline**	(3) **Under 1 SD U5MR** **P2P baseline & HC3 only**
A: Without control variables			
Program impact	0.006	0.114**	0.183**
	(0.028)	(0.049)	(0.075)
Constant	0.003	-0.069*	-0.114**
	(0.019)	(0.035)	(0.043)
B: Without district fixed effects			
Program impact	0.011	0.086	0.143
	(0.036)	(0.063)	(0.088)
Constant	0.001	-0.054	-0.096*
	(0.026)	(0.045)	(0.055)
C: Without ouctome measures aggregated at HC level			
Program impact	-0.001	0.114**	0.170
	(0.029)	(0.056)	(0.158)
Constant	0.004	-0.068	-0.121
	(0.020)	(0.040)	(0.092)
D: Difference between post and pre-treatment values			
Program impact	0.015	0.074	0.099
	(0.032)	(0.055)	(0.079)
Constant	0.030	0.059*	0.020
	(0.021)	(0.031)	(0.044)
Observations (A & B)	7,288	2,860	1,129
Observations (C)	187	73	29
Observations (D)	14,576	5,720	2,258

Notes. The dependent variable is the community monitoring index. The sample in column (1) is the full sample in the full treatment and control arm, in column (2) it is restricted to HCs with a baseline under five mortality rate within one standard deviation of that in Björkman and Svensson (2009), and in column (3) it is further restricted to include HC3s only. Estimates from Equation 1 comparing outcomes between the Full treatment arm and the Control, with the following variations: Panel A shows results without covariates, panel B without district fixed effects, and panel C aggregates outcome measures and covariates at the health center level (the unit of randomization). Panel D shows results from a difference in difference estimation. *** p<0.01; ** p<0.05; * p<0.10

F Replication of Tables in Power to the People

This appendix presents tables replicating Tables 2-6 from Björkman and Svensson (2009) using both our midline and endline data for the full treatment and control group. We use the code provided by Donato and Garcia Mosqueira, who replicated Björkman and Svensson (2009) for 3ie. Their code, in turn, is based on the code provided to them by Björkman and Svensson. We do not have access to the original code. Please see Donato and Garcia Mosqueira (2019) for further details.

Table II: Program impact on monitoring and information

	(1) Suggestion box	(2) Numbered waiting cards	(3) Poster informing services	(4) Poster on patients' rights	(5) Average standardized effect	(6) Discuss facility in LC meetings	(7) Received information about HUMC
Program impact (Midline)	-0.066	-0.000	0.075	-0.029	-0.044	0.005	0.005
	(0.025)	(0.041)	(0.064)	(0.039)	(0.070)	(0.020)	(0.007)
Program impact (Endline)	0.019	-0.005	-0.022	0.057	0.032	0.016	0.014**
	(0.049)	(0.031)	(0.067)	(0.045)	(0.074)	(0.024)	(0.007)
Mean control group (Midline)	0.092	0.111	0.388	0.112	-	0.679	0.064
Mean control group (Endline)	0.234	0.053	0.503	0.168	-	0.498	0.058
Observations (Midline)	187	187	187	187	187	7,204	7,204
Observations (Endline)	187	187	187	187	187	7,288	7,288

Notes. Robust standard errors in parentheses, clustered by catchment area in columns (6)-(7). Columns (1)-(5) are derived from Equation 3 in Björkman and Svensson (2009). Program impact is the coefficient on the assignment to treatment indicator. Outcome measures in columns (1)-(4) are based on data collected through visual checks by enumerators during the endline facility survey; those in columns (6) and (7) are from the endline household survey. All equations include district fixed effects and the following baseline covariates: logarithm of the population density in a 3km radius around the HC, indicator variable for whether the facility has a separate maternity unit, distance to nearest public health provider, number of staff with less than advanced A-level education (collected retrospectively at endline), indicator variable for whether staff have access to a safe source of drinking water, and average monthly supply of quinine. The dependent variables are indicator variables for whether the health facility has (1) a suggestion box for complaints and recommendations, (2) numbered waiting cards for its patients, (3) a poster listing available health services; (4) a poster on patients' rights and obligations; (5) is the average standardized effect of the estimates in columns (1)-(4); (6) is an indicator variable for whether the household discussed the functioning of the health facility at a local council meeting during the past year, and (7) whether the household has received information about the Health Unit Management Committee's roles and responsibilities. *** p<0.01; ** p<0.05; * p<0.10.

Table III: Program impact on treatment practices and management

		Model	Program impact	Midline	Endline	Mean control group	Obs.
(1)	Equipment used (Midline)	DD	0.022	0.019		0.681	14408
			(0.029)	(0.022)			
	Equipment used (Endline)	DD	0.011		0.101***	0.767	14576
			(0.027)		(0.019)		
(2)	Equipment used (Midline)	OLS	0.038*			0.681	7204
			(0.020)				
	Equipment used (Endline)	OLS	0.021			0.767	7288
			(0.014)				
(3)	Waiting time (Midline)	DD	0.541	-16.059***		86	14408
			(4.645)	(2.992)			
	Waiting time (Endline)	DD	-4.695		-24.356***	78	14576
			(4.566)		(2.906)		
(4)	Waiting time (Midline)	OLS	5.701			86	7204
			(3.858)				
	Waiting time (Endline)	OLS	1.015			78	7288
			(3.186)				
(5)	Absence rate (Midline)	OLS	0.023			0.396	187
			(0.030)				
	Absence rate (Endline)	OLS	-0.008			0.405	187
			(0.031)				
(6)	Management of clinic (Midline)	OLS	-0.058			0.101	187
			(0.207)				
	Management of clinic (Endline)	OLS	0.216			0.006	187
			(0.224)				
(7)	Health information (Midline)	OLS	0.020			0.553	7204
			(0.015)				
	Health information (Endline)	OLS	0.030**			0.559	7288
			(0.013)				
(8)	Importance of family planing (Midline)	OLS	-0.007			0.743	7204
			(0.013)				
	Importance of family planing (Endline)	OLS	0.016			0.718	7288
			(0.013)				
(8)	Stockouts (Midline)	OLS	-0.040**			0.122	187
			(0.016)				
	Stockouts (Endline)	OLS	-0.060**			0.263	187
			(0.029)				

Notes. Each row is based on a separate regression. The DD model refers to Equation 2 in Björkman and Svensson (2009), the OLS model to Equation 1. The latter includes district fixed effects and baseline covariates as listed in Table II. Robust standard errors in parentheses, clustered by catchment area in columns (1)–(4) and (7)–(8). Program impact measures the coefficient on the assignment to treatment indicator in the OLS models and the assignment to treatment indicator interacted with an indicator variable for midline/endline in the DD models. The dependent variables are: (1) and (2) indicator variable for whether the staff used any equipment during examination when the patient visited the health facility; (3) and (4) difference between the time the respondent left the facility and the time the citizen arrived at the facility, minus the examination time; (5) share of employed workers not physically present at the time of the endline survey; (6) first component from a principal components analysis of the variables: condition of the floors of the health clinic, condition of the walls, condition of furniture, and smell of the facility, where each condition is ranked from 1 (dirty) to 4 (clean) by the enumerators; (7) indicator variable for whether the household has received information about the importance of visiting the health facility and the danger of self-treatment; (8) indicator variable for whether the household has received information about family planning; (9) share of the last 3 months in which stock cards indicated no availability of drugs (see text for details). *** p<0.01; ** p<0.05; * p<0.10.

Table IV: Program impact on immunization

	(1) Newborn	(2) Under 1 year	(3) 1 year old	(4) 2 years old	(5) 3 years old	(6) 4 years old
Average standardized effect (Midline)	-0.058	-0.072*	-0.098**	-0.023	-0.068*	-0.082*
	(0.078)	(0.043)	(0.041)	(0.034)	(0.041)	(0.048)
Average standardized effect (Endline)	0.046	-0.042	-0.009	0.022	0.005	0.009
	(0.068)	(0.034)	(0.042)	(0.030)	(0.029)	(0.034)
Observations (Midline)	318	1567	1928	2023	2041	1499
Observations (Endline)	356	1537	1584	2042	2170	1722

Notes. Average standardized effects are estimated using Equation 3 in Björkman and Svensson (2009). Dependent variables are indicator variables for whether a child has received at least one dose of measles, DPT, BCG and polio vaccines, and vitamin A supplement, respectively (see text for details); collected at midline and endline, respectively. All models include district fixed effects and baseline covariates, as listed in Table II. Robust standard errors clustered by catchment areas in parentheses. The age brackets refer to children (1) under 3 months; (2) 0-12 months; (3) 13-24 months; (4) 25-36 months; (5) 37-48 months; and (6) 49-60 months. *** $p<0.01$; ** $p<0.05$; * $p<0.10$.

Table V: Program impact on utilization/coverage

	Outpatients	Delivery	Antenatal	Family planning	Average std effect	Use of project facility	Use of self-treatment/ traditional healers	Average std effect
A: Cross-sectional data	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Program impact (Midline)	11.587	0.204	7.105	-4.364	0.005	-0.008	0.012	-0.046**
	(10.425)	(1.645)	(9.468)	(2.751)	(0.072)	(0.005)	(0.005)	(0.020)
Program impact (Endline)	15.649	1.382	18.226	-1.031	0.097	0.001	0.001	0.000
	(11.245)	(1.666)	(8.809)	(3.101)	(0.069)	(0.005)	(0.005)	(0.021)
Observations (Midline)	187	187	187	187	187	7204	7204	7204
Observations (Endline)	187	187	187	187	187	7288	7288	7288
B: Panel data	(9)	(10)			(11)	(12)	(13)	(14)
Program impact (Midline)	7.048	-0.842			0.009	-0.006	0.011	-0.041**
	(8.242)	(1.244)			(0.057)	(0.005)	(0.005)	(0.020)
Program impact (Endline)	11.672	0.286			0.064	0.001	0.001	0.001
	(10.714)	(1.607)			(0.071)	(0.005)	(0.005)	(0.021)
Observations (Midline)	374	374			374	14408	14408	14408
Observations (Endline)	374	374			374	14576	14576	14576
Mean control group (Midline)	220	16	100	30	-	0.4	0.3	-
Mean control group (Endline)	231	14	86	29	-	0.3	0.3	-

Notes. Panel A reports program impact estimates from cross-sectional models with district fixed effects and baseline covariates as listed in Table 2, with robust standard errors in parentheses. Panel B reports program impact estimates from difference-in-differences models with robust standard errors clustered by facility in parentheses. Point estimates, standard errors, and average standardized effects in specifications (1)–(5), (6)–(8), (9)–(11), and (12)–(13) are derived from Equation 3 in Björkman and Svensson (2009). Program impact is the coefficient on the assignment to treatment indicator in the OLS models and the assignment to treatment indicator interacted with an post-treatment indicator variable in the DD models. Dependent variables are monthly averages per health center of (1) number of patients visiting the facility for outpatient care; (2) number of deliveries; (3) number of antenatal visits; (4) number of family planning visits; (5) average standardized effect of estimates in specifications (1)–(4) and (9)–(10), respectively; (6) share of visits to the project facility of all health visits, averaged over the catchment area; (7) share of visits to traditional healers and self-treatment of all health visits, averaged over catchment area; (8) average standardized effect of estimates in specifications (6)–(7) and (12)–(13), respectively, reversing the sign of use of self-treatment/traditional healers. *** $p<0.01$; ** $p<0.05$; * $p<0.10$

Table VI: Program impact on health outcomes (non-corrected U5MR)

	(1)		(2)		(3)		(4)		(5)		(6)	
	Births (Midline)	Births (Endline)	Pregnancies (Midline)	Pregnancies (Endline)	U5MR non-corrected (Midline)	U5MR non-corrected (Endline)	Child death non-corrected (Midline)	Child death non-corrected (Endline)	Weight-for-age z-scores (Midline)		Weight-for-age z-scores (Endline)	
Program impact	-0.001	0.010	0.010	0.013	-4.791	-16.669			-0.088	-0.063	0.074	0.026
	(0.010)	(0.013)	(0.011)	(0.013)	(14.520)	(16.770)			(0.070)	(0.058)	(0.061)	(0.050)
Child age (log)										-1.119***		-1.029***
										(0.037)		(0.038)
Female										0.124**		0.189***
										(0.051)		(0.048)
Program impact x 0 - 1 year old							-0.009	-0.011				
							(0.014)	(0.017)				
Program impact x 1 - 2 years old							0.001	-0.006				
							(0.005)	(0.009)				
Program impact x 2 - 3 years old							0.000	0.001				
							(0.003)	(0.004)				
Program impact x 3 - 4 years old							0.003	0.001				
							(0.003)	(0.004)				
Program impact x 4 - 5 years old							-0.003	-0.005				
							(0.004)	(0.005)				
Mean control group	0.193	0.436	0.340	0.515	92	146	0.020	0.026	-0.510	-0.510	-0.559	-0.559
Observations	7204	7288	7204	7,288	187	187	9,336	12,388	2,264	2,264	2,146	2,146

Notes. Estimates from Equation 1 in Björkman and Svensson (2009), including district fixed effects and baseline covariates as listed in Table II. Non-corrected U5MR refers to the fact that we use reports of children who died in the past 12 months at face-value, as in Björkman and Svensson (2009). Specification (4) also includes a full set of year-of-birth indicators. Robust standard errors in parentheses, clustered by catchment area in columns (1)-(2) and (4)-(6). Program impact is the coefficient on the assignment to treatment indicator. Dependent variables are (1) number of births in the household in the past 12 months; (2) indicator variable for whether any women in the household are or were pregnant in the past 12 months; (3) under-5 mortality rate in the community expressed per 1,000 live births; (4) indicator variable for child death in the past 12 months; (5)-(6) weight-for-age z-scores for children under 18 months excluding observations with recorded weight above the 90th percentile in the growth chart reported in Cortinovis et al. (1997). *** p<0.01; ** p<0.05; * p<0.10

Table VI: Program impact on health outcomes (corrected U5MR)

	(1)		(2)		(3)		(4)		(5)		(6)	
	Births (Midline)	Births (Endline)	Pregnancies (Midline)	Pregnancies (Endline)	U5MR corrected (Midline)	U5MR corrected (Endline)	Child death corrected (Midline)	Child death corrected (Endline)	Weight-for-age z-scores (Midline)		Weight-for-age z-scores (Endline)	
Program impact	-0.001	0.010	0.010	0.013	-3.924	-12.757			-0.088	-0.063	0.074	0.026
	(0.010)	(0.013)	(0.011)	(0.013)	(12.677)	(15.276)			(0.070)	(0.058)	(0.061)	(0.050)
Child age (log)										-1.119***		-1.029***
										(0.037)		(0.038)
Female										0.124**		0.189***
										(0.051)		(0.048)
Program impact							-0.007	-0.008				
x 0 - 1 year old							(0.012)	(0.015)				
Program impact							0.001	-0.005				
x 1 - 2 years old							(0.004)	(0.008)				
Program impact							0.001	0.001				
x 2 - 3 years old							(0.003)	(0.004)				
Program impact							0.002	-0.000				
x 3 - 4 years old							(0.003)	(0.004)				
Program impact							-0.003	-0.005				
x 4 - 5 years old							(0.003)	(0.004)				
Mean control group	0.193	0.436	0.340	0.515	67	125	0.014	0.021	-0.510	-0.510	-0.559	-0.559
Observations	7,204	7,288	7,204	7,288	187	187	9,284	12,336	2,264	2,264	2,146	2,146

Notes. Estimates from Equation 1 in Björkman and Svensson (2009), including district fixed effects and baseline covariates as listed in Table II. Corrected U5MR refers to the fact that we correct reports of children who died in the past 12 months with data on birth and death months collected at endline, thereby ensuring that children who had died more than 12 months prior to the relevant data collection or were not recalled at endline are excluded. Specification (4) also includes a full set of year-of-birth indicators. Robust standard errors in parentheses (3), clustered by catchment area (1)-(2), (4)-(6). Program impact measures the coefficient on the assignment to treatment indicator. Specification (4) also includes a full set of year-of-birth indicators. Robust standard errors in parentheses, clustered by catchment area in columns (1)-(2) and (4)-(6). Program impact is the coefficient on the assignment to treatment indicator. Dependent variables are (1) number of births in the household in the past 12 months; (2) indicator variable for whether any women in the household are or were pregnant in the past 12 months; (3) under-5 mortality rate in the community expressed per 1,000 live births; (4) indicator variable for child death in the past 12 months; (5)-(6) weight-for-age z-scores for children under 18 months excluding observations with recorded weight above the 90th percentile in the growth chart reported in Cortinovis et al. (1997). *** p<0.01; ** p<0.05; * p<0.10

G Comparison of ACT Health and Power to the People

	Difference	P2P	ACT Health
Program Design & Implementation	1. Intervention dates	2004-2005	2014-2016
	2. Number of program variations	1. Control 2. Full intervention	1. Control 2. Information and mobilization only 3. Interface meeting only 4. Full intervention
	5. Involvement of community-based organizations (CBOs)	Worked through 18 CBOs with prior experience in some of the treatment communities	Worked in consortium with 4 implementing partner organizations with no prior experience in the treatment communities
	6. Length of time of community dialogue meeting	2 half-days	1 half-day
	7. Role-playing activity?	Yes	No
	8. Avg. number of participants at community dialogue meetings	150	100
Research Design & Data Collection	9. Districts included in study	9	16
	10. Interval between intervention and collection of outcome data	~1 year	8 months (to midline) 20 months (to endline)
	11. HC sample inclusion criteria	HC3s only	HC2s and HC3s
	12. HC catchment area definition	5km radius around each HC	Nearest three villages to each HC
	13. HC sample size	50 HCs (only HC3s)	Total: 376 HCs (225 HC2s and 157 HC3s) Full treatment & control: 189 HCs
	14. Household sample size	5,000 households; 100 per HC catchment area	14,609 households (7,288 households in full treatment & control); ~40 per HC catchment area

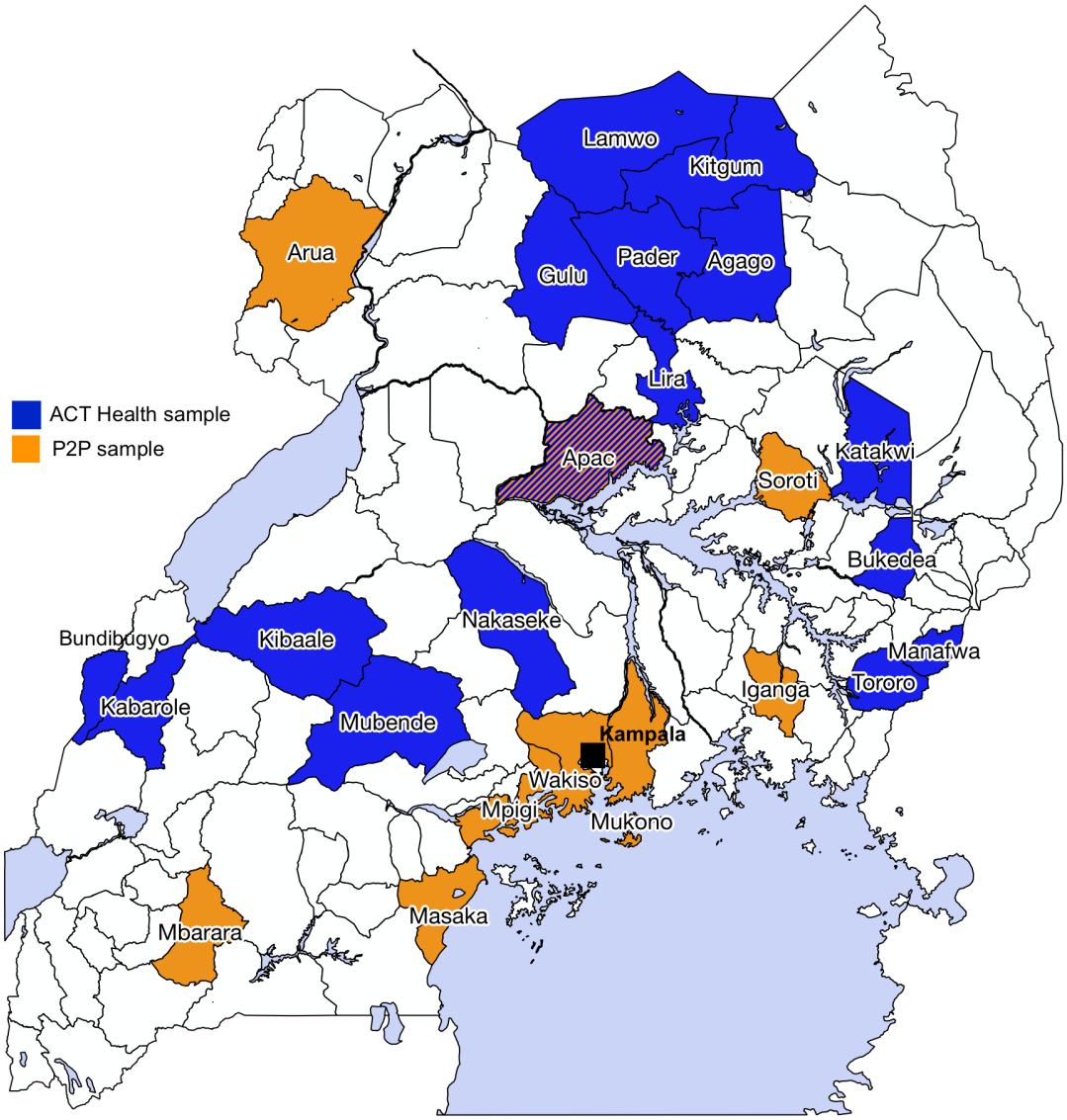


Figure 7: Map indicating the districts in the ACT Health and P2P samples

Table G1: Baseline comparison – P2P vs. ACT Health

	(1) P2P	(2) ACT Health Full sample	(3) ACT Health HC3 only	(4) ACT Health HCs w/in 1 SD	(5) ACT Health (3) and (4)
A. Utilization					
Number of outpatients in HC (HMIS)	674.733 (286.213)	200.073 (93.220)	268.806 (89.074)	194.787 (93.685)	252.216 (100.063)
Number of deliveries at HC (HMIS)	7.480 (6.798)	13.841 (14.823)	21.349 (15.818)	13.286 (14.450)	20.953 (14.857)
Number of visits to HC	3.184 (5.583)	14.186 (12.153)	14.217 (12.344)	14.078 (12.098)	12.637 (11.633)
% of visits to HC, vs. other providers	0.304 (0.335)	0.377 (0.238)	0.374 (0.231)	0.360 (0.228)	0.341 (0.230)
% of visits to traditional healers or self-treatment, vs. other providers	0.334 (0.336)	0.311 (0.210)	0.318 (0.206)	0.309 (0.215)	0.319 (0.220)
Vaccination rates, children <36 months	0.066 (0.249)	0.763 (0.425)	0.766 (0.424)	0.755 (0.430)	0.752 (0.433)
B. Treatment quality					
Used equipment	0.481 (0.500)	0.664 (0.418)	0.695 (0.414)	0.694 (0.402)	0.762 (0.365)
Waiting time	143.412 (98.379)	101.501 (81.441)	118.171 (88.232)	104.764 (80.568)	116.321 (87.437)
% staff presence	0.534 (0.265)	0.268 (0.198)	0.222 (0.163)	0.266 (0.183)	0.191 (0.154)
Stockouts	0.496 (0.208)	0.086 (0.158)	0.077 (0.132)	0.064 (0.128)	0.060 (0.098)
C. Patient satisfaction					
Polite staff	0.910 (0.286)	0.906 (0.258)	0.896 (0.273)	0.913 (0.245)	0.910 (0.245)
Staff interested in health	0.905 (0.293)	0.907 (0.255)	0.901 (0.266)	0.911 (0.244)	0.906 (0.251)
Free to express clearly	0.946 (0.226)	0.833 (0.322)	0.825 (0.333)	0.827 (0.321)	0.811 (0.331)
D. Health outcomes					
Weight-for-age	0.789 (0.610)	0.810 (0.776)	0.826 (0.763)	0.810 (0.774)	0.838 (0.798)
U5MR	144.400 (94.446)	12.771 (41.748)	15.628 (48.237)	35.046 (63.686)	40.930 (72.077)

Notes. The table displays the mean of each variable in the control group at baseline, with standard deviations in parentheses, across different samples: column (1) shows descriptive statistics for the control group of Björkman and Svensson (2009), column (2) for the pure control group in ACT Health, column (3) for the subset of (2) consisting of HC3s, column (4) for the subset of (2) consisting of health centers with a baseline under 5 mortality rate within one standard deviation of that in Björkman and Svensson (2009), and column (5) for the the intersection of (3) and (4). For variables indicated with $\diamond$, baseline measures are not available for P2P, so column (1) shows the mean in the control group at endline instead. All data is from household surveys, except for the top two variables, which are sourced from administrative health center data (HMIS), and data on staff presence and drug stockouts, which are sourced from the health center survey.

H Implementation of ACT Health

H.1 Implementing Organizations

The Irish NGO, GOAL, launched the ACT Health program in 2014 with funding from DFID. The program was implemented by three Ugandan regional partners across 15 districts and by GOAL Uganda in one district. All three regional partners had prior experience working on community mobilization and/or public health, and already had a strong footprint in the region, if not the districts, they implemented ACT Health in. In particular, the Coalition for Health Promotion and Social Development (HEPS), which was founded in 2000, focused on access to health care and essential medicines, maternal health rights, community-based empowerment work, and health advocacy prior to becoming involved in the implementation of ACT Health. The Multi-Community Based Development Initiative (MUCOBADI) was started by HIV positive teachers in 2000 and focused on HIV prevention, access to primary health care, community mobilization, and livelihood development. Finally, the Kabarole Research and Resource Centre (KRC), which was founded in 1996, focused on leadership mobilization, good governance, and research and advocacy. Two of the three regional partners, HEPS and MUCOBADI, were actively involved in the design and implementation of a pilot of the ACT Health program in Bugiri district. Organizational leadership aside, implementing staff for the intervention were specifically recruited for the program and had to have prior experience in community mobilization and/or public health. All implementing staff underwent extensive training and were continuously monitored and supervised by GOAL Uganda.

H.2 Implementation Monitoring

GOAL ensured fidelity to both the intervention and the randomized impact evaluation protocol through several quality control measures.

- Detailed procedure manuals for each of the three variations of the program (the full program, the information and mobilization program—called separate dialogues by GOAL—and the interface-only program). These manuals were used for training and as a reference to ensure that all partners had clear and precise instructions regarding every detail of the intervention. These manuals were used for training and as a reference so that all implementing partners had clear and precise instructions of every detail of the interventions.
- Extensive monitoring data captured in an online monitoring database, which tracked the dates and numbers of people participating in each dialogue. The database also includes the actions agreed upon in the action plans and social contracts developed during the dialogues, and tracked their progress at each follow-up meeting. The reports include data on the dates and number of people participating for each program activity, including HC and community meetings, interface meetings, and each of the follow-up meetings. They also record all the actions agreed to in the action plans and social contracts and track their progress at the follow-up meetings.
- Direct observation by GOAL's monitoring team. To assure quality across the life of the intervention, GOAL had "mentor" managers and monitoring, evaluation, and learning (MEL)

officers embedded within the teams of each partner organization to provide direct support and programmatic guidance, as needed. Mentor manager and MEL officers observed a portion of each partner's dialogues and follow-ups for quality assurance purposes. In particular, 97% of health centers were monitored at least once by either a mentor manager or MEL officer (see Section E of Table H1 for more details). During these direct observations, officers recorded information about facilitator behavior, the presentation of the citizen report cards, the nature of participation during the meeting, and whether the action plans and social contracts met certain quality criteria. The feedback tool for these observations is reproduced below.

- Issues tracking. GOAL tracked issues as they came up during implementation. They gave field teams a detailed protocol of issues to watch out for and flag. Issues were shared with the evaluation team and solutions were jointly decided to resolve the issues.

Table H1: GOAL attendance and monitoring data

	Full treatment			Information only			Interface only		
	N	**Mean**	**SD**	**N**	**Mean**	**SD**	**N**	**Mean**	**SD**
A. HC information separate dialogue									
# of HC staff present	90	6.14	3.26	92	6.24	3.42	n/a	n/a	n/a
% of HC staff present	90	67.25	33.89	92	71.04	27.56	n/a	n/a	n/a
B. Community information separate dialogue									
# of community members present	90	102.36	20.99	91	99.01	25.17	n/a	n/a	n/a
# of female community members present	90	35.67	8.79	91	34.26	9.91	n/a	n/a	n/a
% of HCs where at least one SC official is present	90	0.10	0.30	91	0.36	0.48	n/a	n/a	n/a
C. Interface meeting									
# of community members present	91	33.10	14.94	n/a	n/a	n/a	97	67.38	13.67
# of female community members present	90	16.87	8.83	n/a	n/a	n/a	97	34.79	10.49
# of HC staff present	91	3.97	2.31	n/a	n/a	n/a	97	3.80	2.45
% of HCs where at least one SC official is present	91	0.30	0.46	n/a	n/a	n/a	97	0.35	0.48
D. Follow-up meetings (average across three meetings)									
# of community members present	91	41.01	14.77	92	54.51	20.14	97	55.67	17.54
# of female community members present	91	20.54	8.13	92	27.37	11.01	97	29.21	11.41
# of HC staff present	91	3.71	2.22	90	3.14	3.30	97	3.46	2.10
% of HCs where at least one SC official is present	91	0.39	0.54	.	.	.	97	0.47	0.60
E. Monitoring & oversight									
% supervision during the initial activity	92	0.80	0.40	92	0.62	0.49	97	0.58	0.50
% supervision during at least one follow-up meeting	92	0.76	0.43	92	0.92	0.27	97	0.73	0.45
% supervision at least one time (initial activity or follow-up)	92	0.98	0.15	92	0.99	0.10	97	0.93	0.26

Notes. Data is drawn from implementer's monitoring tools, and verified by GOAL's monitoring team through direct observation in the share of meetings indicated in Panel E.

Version 10-Mar-2015

Observation-Feedback Tool for ACT Health Community Activities

*This form is for use by Line Managers and Mentorship Managers when observing **dialogues** and **interfaces.** Compete for each observation and share your feedback with the Officers after the visit. Copy should be kept in file of Officer and Observer.*

Observerer	Name			Title		Organisation	
Date of observation	Day	Month	Year	**Procedure # (Tick One ONLY)**	#2	#3	#4
Location	Health Centre			Sub-county		District	
Activity observed	h of WDWW. Linking the HC Dialogue		Community Dialogue	Interface			
Facilitators Observed	Name		Organisation	Name		Organsiation	

	Was the Facilitator . . .	YES	NO	Partly
Facilitator behaviour	Well-prepared on the process and purpose of meeting?			
	Effective in managing expectations of participants?			
	Clear and audible so participants understood?			
	Managing time? (Covered all agenda items, not exceed 5 hrs)			
	Listening actively? (Eye contact, summarising, repeating)			
	Enabling diverse participants to speak and participate?			
	Ensuring mutual respect among participants?			
	Bringing focus to Responsibility of community members?			
	Bringing focus to Responsiveness of HC staff?			
	Impartial? (non-biased, non-judgemental, not taking sides)			
	Managing challenges? (conflicts, disturbances)			
CRC Presentation **Applicable to** Separate Dialogues Full Programme Observer must also study the CRC before meeting	**Did the Facilitator(s)…**	**YES**	**NO**	**Partly**
	Clearly understand the CRC content before the meeting?			
	Know and clearly communicate the MoH standards?			
	Bring the appropriate posters for the HC level?			
	Present clearly? (word choice, local language, explain boxes)			
	Check to ensure that participants understood the CRC?			
	Use CRC information to challenge perceptions and excuses?			
	Use probing questions to improve Responsibility?			
	Use probing questions to improve Responsiveness?			
	Use probing questions to improve Relationships?			
	Ensure mis-conceptions raised by participants are corrected			
Participation	**How participatory was the meeting?**	**YES**	**NO**	**Partly**
	Were *all* HC staff present?			
	Were all social groups represented as per mobilisation list?			
	Were all small groups working effectively?			
	Was the participation in the large group balanced? Women talked?			
	Was the posture and positioning of facilitator conducive?			
Action Plan/Social Contract	**Action plan/social contract meet basic quality criteria**	**YES**	**NO**	**Partly**
	Issues are clearly stated			
	Issues are related to the CRC/Information in CRC (Procedure #2 & #4)			
	Actions are related to the issue			
	Actions are achievable with local resources (low/no cost)			
	Inclusive of actions for community AND health centre staff			
	How many priority issues from women's groups are included?			
	Inclusive of issues/actions from all social groups?			
	Mix of actions for now (6months), soon (12 months) and later			

Observation-Feedback Tool for ACT Health Community Activities

*This form is for use by Line Managers and Mentorship Managers when observing **dialogues** and **interfaces.** Compete for each observation and share your feedback with the Officers after the visit. Copy should be kept in file of Officer and Observer.*

Observerer								
	Name			Title			Organisation	
Date of observation				**Procedure # (Tick One ONLY)**				
	Day	Month	Year		#2	#3	#4	
Location								
	Health Centre			Sub-county			District	
Activity observed	h of WDWW. Linking the							
	HC Dialogue		Community Dialogue		Interface			
Facilitators Observed								
	Name		Organisation		Name		Organsiation	

Feedback for Facilitator/Officer

The ACT Health programme has a component of on-the-job training and support. This support (mentorship) is very important for continuous learning and implementation of a high quality programme. These feedback tips should be shared with the faciltiators(s) after the meeting or in a visit to the Officer shortly (within three (3) working days) of the observation.

WELL DONE! These are the areas where you excelled. Thanks for your work!	
1)	
2)	
3)	
There are a few things you can work on for next time. Let me know how I can support you best.	
1)	
2)	
3)	

As an observer, what do you think participants in this community feel about the ACT Health programme?

Observer	
Name	
Signature	
Date of sharing with facilitator	

Facilitator	
Name	
Signature	
Date of discussing with facilitator	

H.3 Intervention Materials

H.3.1 Steps of the intervention

Procedures Table

Procedure		Citizen Report Card (CRC)	Health Centre Dialogue	Community Dialogue	Interface	Final Output Action Plan or Social Contract	Follow-up every six (6) months
1	No intervention (control)	Baseline data will be collected but no activities.	None will be held.	None will be held.	None will be held.	None will be developed.	*Survey at 12 months and 36 months after baseline.
2	Information provided (CRCs) and two separate action plans are developed in community and health centre dialogues. There is no interface between HC staff and community.	The Citizens Report Card will be shared in community dialogue and health centre dialogue.	Health centre staff have a dialogue and develop an action plan.	Community members have a dialogue and develop an action plan.	This will not be held.	Two separate action plans will be developed – one by the health centre staff and one by the community members.	Every six months separate follow-up dialogues for community members and health centre staff. *Survey at 12 months and 36 months after baseline.
3	No information (CRC) provided and no health centre or community dialogues are held. Interface between health centre staff and communities yields one social contract.	The Citizens Report Card will not be shared.	This will not be held.	This will not be held.	The interface will bring together community members and health centre staff.	One social contract developed at the interface combining community and HC actions.	Every six months follow-up interface with community members and HC staff. *Survey at 12 months and 36 months after baseline.
4	**Information (CRC) provided in separate dialogues at health centre and community levels. During these dialogues, each group develops an action plan which is discussed at the interface. The interface yields one social contract.**	**The Citizens Report Card will be shared at community dialogue and health centre dialogue.**	**Health centre staff dialogue and develop an action plan.**	**Community dialogue and develop an action plan.**	**The interface will bring together health centre staff and representatives of the communities.**	**One social contract developed combining community and HC actions.**	**Every six months follow-up interface with community members and HC staff jointly. *Survey at 12 months and 36 months after baseline.**

ACT Health Procedure #4 Full Programme

6

H.3.2 Example of a mobilization protocol and citizen report card

Version: 27-August-2014

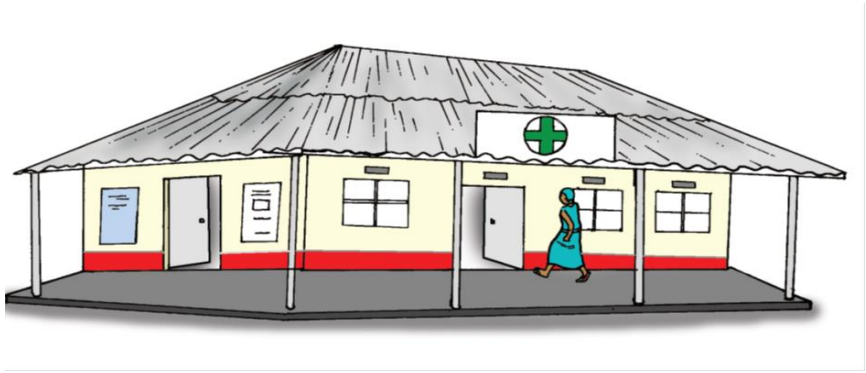
Community Mobilisation List

Mobilising diverse social groups (women and men of different ages, income levels and social standing) is very important! We want to hear voices of all social groups in the community. Please think of someone in the village who meets the social group description below and would be willing to participate. Thank you for your time!

Health Centre	
Village	
Name of VHT carrying out mobilisation	

Dialogue Meeting Participants
*at least 50% of the participants from each village should be women

Social Group to Target	Individual Name	Will attend interface meeting?
1. LC1 Chairperson		
2. LC Women Representative		
3. LC Youth Representative (15-20 years old)		
4. LC Representative with disability		
5. Mother		
6. Mother		
7. Mother		
8. Male Lowest Income Group		
9. Female Lowest Income Group		
10. Female youth (15-20) in Lowest Income Group		
11. Male youth (15-20) in Lowest Income Group		
12. Male Highest Income Group		
13. Female Highest Income Group		
14. Male youth (15-20)		
15. Female youth (15-20)		
16. Male adult (21-49)		
17. Female adult (21-49)		
18. Male elder (50+)		
19. Female elder (50+)		
20. VHT Member		
21. VHT Member		
22. VHT Member		
23. VHT Member		
24. VHT Member		
25. HUMC Member		



Kagote Health Center III

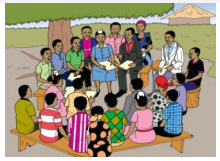
Kabarole District

Citizen Report Card

Survey dates: 10th October 2014 to 26th October 2014

Report Card Prepared: 27th October 2014

Responsibility	Responsiveness	Relationships
Individuals have good health-seeking behaviour. They seek preventive care (ANC, immunisations, testing, etc.) and go early for treatment of illness to avoid complications.	Health Center staff use resources effectively and provide care as per Ministry of Health standards in the Uganda National Minimum Health Care Package (UNMHCP).	Mutual understanding and trust between community members and health Center staff. Includes better understanding of each other's constraints.



Note: This Citizens Report Card has been compiled from responses to household surveys and HC staff interviews.

Rights and Responsibilities

Issue	Households say	Health Center says
Who could name at least 5 health rights and entitlements	0%	Could name 3

Health Rights

Some Major Health Rights in Uganda Patient's Charter
Right to choose
Right to complaint and redress
Right to access essential medicine
Right to access information
Right to privacy and confidentiality

Health Responsibilities

Health Responsibilities include
Responsibility to be healthy
Responsibility to participate

What services does our HC III provide?

GOVERNMENT STANDARD* Services that should be provided by HCIII	Health Center says
Antenatal care	Yes
Delivery	No
Outpatient care	Yes
HIV counselling and testing (HCT)	Yes
Immunisation	Yes
Lab services	Yes
Family planning methods (simple)	Yes
Family planning methods (advanced)	Yes
Health education (at HC)	Yes
Family planning education	Yes
Health Outreach (villages)	Yes
Prevention of Mother to Child Transmission (PMTCT)	Yes
Anti-retroviral therapy (ART)	Yes

*Uganda National Minimum Health Care Package

How many people use this HC? (Responsibility)

The community member visits to Kagote health Center III in the past 12 months.

Use patterns (adults and children)
17% of all health care visits in this community were to Kagote health Center

Reasons why community do not use health Center	
Reasons why the households in the community DO NOT visit Kagote health Center	60% Lack of drugs 20% Long waiting time 20% Long Distance 0% Cannot afford payment 20% Attitude of Staff 0% Unclean facility 10% Poor quality services 10% Have not been sick 10% Don't provide treatment I need 10% Others

Community member visits to other health providers in the past 12 months.

Other providers	Average utilisation (adults and children)
Private not for Profit (PNFP) e.g. NGO, missionary health Center	5% of all health care visits
Private for profit	19% of all health care visits
Traditional healer	1% of all health care visits
Community health worker e.g. VHT	6% of all health care visits
Self-treatment (pharmacy, drug shop)	32% of all health care visits
Other government health facilities e.g. HC III, IV, hospital	20% of all health care visits

How does our community compare?		
Health care provider	Kagote health Center	District use patterns of nearest government health centers
Use patterns	17%	31%

How many of us use ANC and deliver at our HC III? (Responsibility)

GOVERNMENT STANDARD = pregnant mothers should have four (4) ANC visits

Community's utilisation of antenatal care, family planning	
Percentage of households with pregnant women who have visited **Kagote health Center** for **antenatal care** since **September 2013**	63%
Percentage of those pregnant in the last year who **delivered** at **Kagote** health Center since **September 2013**	0%
Percentage of women who received an **HIV test** during ANC visit **(PMTCT)**	100%

Reasons why we (community members) do not deliver at this HC

Why do pregnant women in the community choose **NOT** to deliver at **Kagote** health center	**0%** Cannot afford **20%** Health Center was not open **0%** Use traditional birth attendant **0%** Attitude of staff **0%** Was not treated well at the HC **10%** Delivered quickly **20%** Referred to another health center **20%** Other provide better services **20%** Did not have the requirements **40%** Other

How do we compare? Antenatal care and maternity care		
Use pattern of antenatal care and maternity care	Among pregnant women in this community	Among pregnant women in Kabarole District
Percentage of households with pregnant women who have visited **their closest government** health Center for antenatal care	**63%**	**46%**
Percentage of pregnant women who made four (4) ANC visits to the nearest health center.	**10%**	**23%**

How do we compare? Immunisation		
Immunisation	In this community	*Among children in District*
% of children <5 immunised in **Kagote** catchment area	**98%**	**98%**

How many of us use family planning services at our HC III? (Responsibility)

Community's utilisation of family planning	
Percentage of households who have visited **Kagote health Center** for **family planning** since **September 2013**	29%

Why do households in the community choose NOT to use family planning services at Kagote health Center?	**0%** Attitude of staff **9%** Use natural methods **N/A** Not interested **Data not collected** **6%** Fear side effects **3%** Do not provide family planning education **19%** Do not need (young/want children/too old) **0%** Partner does not want **19%** Go elsewhere **3%** Health center lacks family planning drugs **3%** Did not know about the service **3%** Refused to answer **41%** Other

What community says about staff attendance at our HC III (Responsiveness)

GOVERNMENT STANDARD = absenteeism is any *unexcused* absence

Percentage of household saying medical staff attend work at Kagote health Center	
Always at work	71%
Sometimes at work	22%
Rarely at work	7%

GOVERNMENT STANDARD = HC III should have eleven (11) medical staff + eight (8) other staff for a total of nineteen (19) staff

Type of Staff	Government Standard	Staff actually allocated	Staff present on survey day
Medical	11	12	8
All staff	19	18	11

Medical staff attendance at Kagote health Center on survey day	
Total number of medical staff out on leave and/ or training on the survey day	2
Total number of medical staff out for outreach on the survey day	0
Percentage of households who said the health Center was open when they last visited	93%

What community says about drug availability

Household rating of drug availability

Household rating of drug availability at Kagote health Center	
Patients who received drugs at their last visit	88%
Drugs are always available	16%
Drugs are sometimes available	76%
Drugs are rarely available	9%

Do community members know when drugs are received?

Health issue	Households say	Health Center says
Do you know when drugs are delivered to Kagote health Center?	31% yes	Yes, we do distribute information on drug deliveries

Households reporting about the drugs they have	
Average number of type/brands of drugs received per visit per person	2
Percentage of patients who say it was clearly explained how to take the drugs	79%

GOVERNMENT STANDARD = All six (6) items should be available at all times

Health Center reporting stock outs of the following tracer items in the last 3 months	
1. Cotrimoxazole (CTX)	No
2. Artemether/Lumefantrine	No
3. Oral Rehydration salts (ORS)	Yes
4. Depo Provera	No
5. Measles Vaccine	No
6. Sulfadoxine and Pyrimethamine (SP)	No

Minimum standard drug storage conditions	
Method in place to control temperature	Yes
Windows that can be opened or there are air vents	Yes
Direct sunlight cannot enter the area	Yes
Area is free from moisture	Yes
Cold storage in the health Center	Yes
Medicines are stored directly on the floor	No
There is evidence of pests in the area	No

Fees at our HC (Responsiveness)

Government Standard	Health Center says	Community says	
0.00 UGX for government health facilities	No	Average amount paid	
		Cash	Value: In kind
		300 UGX	0 UGX

What did we bring / buy most?	
Top 3 things that have to be bought or brought to Kagote health Center	1.Exrecise book for prescription 2.N/A 3.N/A

Fees – HC III services	Households say	Health Center says	District Averages (Households say)
User fees (Cash)	2%	No	2%
Average amount paid for user fees (cash)	300 UGX	N/A	3,610 UGX
User fees (In-kind)	0%	N/A	0.1%
Average amount paid for user fees (in-kind)	N/A	N/A	1,200 UGX
Center charges for antenatal care (answered by pregnant women)	10%	No	1%
Average amount paid for antenatal care	300 UGX	0 UGX	1,650 UGX
Center charges for delivery (answered by women who delivered there)	0%	No	2%
Average amount paid for delivery	N/A	0 UGX	5,000 UGX
Center charges for drugs (including injections)	0%	No	0.1%
Average amount paid for drugs	N/A	0 UGX	1,800 UGX
Patients have to pay for immunization	0%	No	0.4%
Average amount paid for immunization	N/A	0 UGX	2,890 UGX

Satisfaction

Waiting times

GOVERNMENT STANDARD = waiting time should be less than one hour

Waiting time until first attended to

Government Standard	Community says	Health Center says
Less than 1 hour	00 Hour 39 Minutes	30 Minutes

Health Unit Management Committees (HUMCs)

Percentage of households who DO know at least two (2) roles of the HUMC	4%

Satisfaction with Relationships between HC Staff and Community

Overall satisfaction with relationship between community members and HC staff	Households say	HC says
Very satisfied	18%	Satisfied
Satisfied	60%	
Not satisfied	22%	

Health issue	Households say	Health Center says
Were patients treated politely	86% yes they were polite/ extremely polite	" Yes, we sometimes treat patients politely"
Average exam time for patients at their last visit	14 minutes	45 minutes

Health issue	Percentage
Percentage of patients who said the health worker listened to what they said at their last visit	89% yes/ very interested and asked questions
Percentage of patients who said the staff clearly explained their medical condition	70%
Percentage of patients who were examined at their last visit	65%
Percentage of patients who said health worker wore uniform at their last visit	59%
Percentage of patients who said they had privacy during the examination at their last visit	89%

How do we compare?		
Household says	Kagote health Center	District
Average waiting time for patients	39 minutes	46 minutes

H.3.3 Example of a community action plan, health center action plan, and social contract

Instructions

Please record below the action plan that was developed. Please do not edit. Type it exactly as members developed. You will need a verison in the local language and you will work with the secretary to translate to English for analysis/tracking.

For **Procedure #2 (Separate Dialogues)** PLEASE USE THE SOCIAL CONTRACT TEMPLATE Because the actions in the social contract template are final, not "suggested."

For **Procedure #3 (Interface Only)** and **Procedure #4 (Full Programme)**, the action plan should be placed in the file. Only the social contract is submitted with the report to your manager.

Document Type	**ACTION PLAN**			For Document type enter "Community Action Plan" or "HC Action Plan"
District	**Kabarole**			
Sub-County	**West division**			
Health Centre	**Kagote health center III**			
Procedure #	**4**			
Facilitator name(s)	**Makasi and Hilary**			
Facilitator Organisation	**KRC**			
Action Plan By (tick one)				
	COMMUNITY			
Date developed	1	12	2014	
	Day	Month	Year	

#	Issue	Reasons for Issue	***Suggested*** **Action**	***Suggested*** **Person Responsible**	***Suggested*** **Completion Date**
1	Information on drug availability	Community is not informed whenever drugs are available at the health center	Writing to the in charge to always inform the community through the notice board whenever drugs are available at the health center	VHT Nkayezu	30/12/2015
2	Staff putting on uniform	Not following up staff who come on duty without putting on uniform by the incharge	Writing to the in charge to speak to the staff to always put on Uniform while on duty	VHT Coordinator Richard mwagushia	30/1/2015
3	Lab equipments	Some of the Lab equipment's are not at the health center like the one for Typhoid	Writing to the in charge to ensure that all the Lab equipment's are available at the health center	VHT Nkayezu	30/1/2015
4	The behavior for the staff	The staff do not mind about the patients at the health center	Writing to the in charge to speak to the staff about there behaviors in handling the patients at the health center and come up with the solution	Kairu christopher (Elder)	30/1/2015
5	UHMC roles	Community does not know the role of HUMC at the health center	Writing to the incharge to inform the community on the role of HUMC through the village notice boards	incharge Mugisa Brian	30/1/2015
6	Coming late by patients for treatment at health center	Community members have a thinking that there are always no drugs at the health center	HVTs should sensitize the community members to always come early for treatment at the health center before the sickness worsens	VHT Coordinator Richard mwagushia	30/12/2015

Figure 8: Sample Community Action Plan from Kabarole-Kagote HC3

Instructions
Please record below the action plan that was developed. Please do not edit. Type it exactly as members developed. You will need a verison in the local language and you will work with the secretary to translate to English for analysis/tracking.
For **Procedure #2 (Separate Dialogues)** PLEASE USE THE SOCIAL CONTRACT TEMPLATE Because the actions in the social contract template are final, not "suggested."
For **Procedure #3 (Interface Only)** and **Procedure #4 (Full Programme)**, the action plan should be placed in the file. Only the social contract is submitted with the report to your manager.

Document Type	**ACTION PLAN**			For Document type enter "Community Action Plan" or "HC Action Plan"
District	**KABAROLE**			
Sub-County	**WEST DIVISION**			
Health Centre	**KAGOTE HEALTH CENTER 3**			
Procedure #	**4**			
Facilitator name(s)	**MAKASI & HILARY**			
Facilitator Organisation	**KRC**			
Action Plan By (tick one)				
		HEALTH CENTRE		
Date developed	24	11	2014	
	Day	Month	Year	

#	Issue	Reasons for Issue	***Suggested*** **Action**	***Suggested*** **Person Responsible**	***Suggested*** **Completion Date**
1	Bringing mothers to deliver at the health center	Service was not being offered at the health center	Using VHTs to give information that the center now conducts deliveries, Carry out health education at the HC	Health assistant Muhumuza Michael	end of Feb 2015
2	Information on drug availabiltiy	Information gap	Displaying dlivery of drug on public notice boards	Medical records officer Henry & Beatrace	end of Jan 2015
3	Information on health rights & responsibility	Information gap	Display alist of health rights and responsibility on the public notice boards	In cahrge Mugisa	end of Feb 2015
4	Involvement of HUMC in HC activity	Not community members & have over stayed in office	Write to the office of the town clack about formation of HUMC at the HC	In charge	end of march 2015
5	Community dialogue	CRC not disemineted to the community	Disemineting the CRC	Hilary and Makasi	end of Dec 2014

Figure 9: Sample HC Action Plan from Kabarole-Kagote HC3

Instructions

Please record below the social contract that was developed in the Interface. Please do not edit - type it exactly as members have developed. You will need a verison in the local language and you will work with the secretary to translate to English for analysis/tracking.

For **procedure #2 (Separate Dialogues)** please use this format to develop action plans.

For **Procedure #3 (Interface Only)** and **Procedure #4 (Full Programme)**, attach a copy of the social contract to the Interface report and submit to Manager within five (5) working days of the Interface.

Document Type	**SOCIAL CONTRACT**		
District	KABAROLE		
Sub-county	WEST DIVISION		
Health Centre	KAGOTE HEALTH CENTER 3		
Procedure #	4		
Facilitator name(s)	MAKASI K. EDWARD & RUYOOKA HILARY		
Facilitator Organisation	KRC		
Date developed	8	12	2014
	Day	Month	Year

For "Document" you will enter "social contract" if this is used in Interface. Enter "Commuity Action Plan" or "HC Action Plan" for Procedure #2 (Separate Dialogues)

#	Issue	Action	Person Responsible	Expected Completion Date	Evidence of Progress on Action	Person Responsible for Monitoring Progress
1	Putting on Unifrom by staff	Incharge to inform the staff in a meeting to always put on uniform while on duty at the health center	In-charge Mugusa Bra	End of March 2015	Finding staff in uniform while on duty and the minuts for the meeting	O/c Kagote station Muhindo
2	Staff conduct	Incharge to hold a meeting with staff to discuss about there conduct towards the patients while on duty at the health center and come up with a solution	In-charge Mugusa Bra	End of Feb 2015	Minutes for the meeting and the change in the conduct for the staff towards the patients	Kabasiguzi Beatrace (elder)
3	HUMC Functionality	Electing the new HUMC	Town clark and the incharge	End of June 2015	Council Minutes and the new HUMC to be in place	Tuhaise Aisha (elder)

Figure 10: Sample Social Contract from Kabarole-Kagote HC3