Community Health Provision at Scale:

Evidence from a Cluster-Randomized Trial in Uganda

Phyllis Awor[i], Martina Björkman Nyqvist[ii], Andrea Guariso[iii], Jakob Svensson[iv]

[i] Makerere University, Kampala, Uganda

[ii] Stockholm School of Economics, Stockholm, Sweden

[ii] University of Milan-Bicocca, Milan, Italy

[iv] IIES, Stockholm University, Stockholm, Sweden

Corresponding author: Jakob Svensson, IIES, Stockholm University, 106 91 Stockholm, Sweden. Tel: +46 8 163060, Fax: +46 8 161443, jakob.svensson@iies.su.se

Abstract

Background Evaluations of proof-of-concept studies on community health worker (CHW) programs suggest that well-designed and well-managed CHW initiatives can significantly improve health services and reduce child morbidity and mortality. However, evidence on the sustainability of these effects at scale remains limited. We assessed the impact of a CHW program that had previously been shown to reduce under-five mortality in Uganda. The trial incorporated elements of an adaptive experimental design: performance outcomes were collected and shared with two program implementers midway through the initial trial period. This allowed for evidence-based adjustments to the intervention's design and implementation guidelines, as well as modifications to the trial duration, while maintaining the original treatment assignments.

Methods A cluster-randomized controlled trial was embedded within the nationwide scale-up of a CHW program. The same program was implemented by two NGOs operating in different regions of the country in partnership with the Ministry of Health across 500 clusters in 13 districts in Uganda. Clusters were stratified by implementer and across 15 geographical zones. In all intervention clusters, financially incentivized CHWs were deployed over an initial three-year period, which was later extended to five years (2016–2021). The study was conducted in two phases: Phase I spanned from trial initiation to the midline performance evaluation, while Phase II began when implementers received the evaluation findings and continued for 28 months until the extended evaluation concluded. Each cluster had an average of 20 surveyed households, with a total of 10,172 households followed by the trial's end. The primary outcome was all-cause child mortality.

Findings A performance review conducted approximately one year into the trial showed only a minor increase in household interactions with CHWs and no consistent improvements in service provision. In response, NGO A introduced major modifications to the design and implementation of the program (henceforth Program A), whereas NGO B introduced only minor adjustments to it (henceforth Program B). The trial period was extended by two years to assess the potential benefits of these revisions. Across the full study period, under-five mortality did not significantly differ between intervention and control clusters. However, during Phase II, among children exposed to Program A, child mortality declined by 28% (adjusted RR 0.72, 95% CI: 0.54–0.97). In contrast, no significant reduction in child mortality was observed for Program B in Phase II. Secondary outcomes improved significantly for households in Program A, while Program B showed mostly modest gains, consistent with the mortality findings. No harm was reported.

Conclusion How to ensure that CHW programs remain effective at scale remains an open question. Our findings suggest that an adaptive program approach, combined with implementers who are both willing and able to make evidence-based design adjustments, can help sustain the impact of successful pilot programs.

Funding Children Investment Fund Foundation. The trial was registered in the Pan African Clinical Trials Registry (PACTR201609001398349) and in the American Economic Association RCT registry (AEARCTR-0002392).

Introduction

The World Health Organization (WHO) promotes the deployment of Community Health Workers (CHWs) as a key strategy to achieve Universal Health Coverage and improve child survival in low- and middle-income countries (LMICs). [1-3] Four proof-of-concept studies on CHW programs have demonstrated significant reductions in neonatal and child mortality.[4-7] However, these studies are often confined to specific areas with a limited number of clusters and benefit from substantial supervision and support from implementing organizations and other agencies. Scaling these programs nationally presents challenges in maintaining rigorous monitoring and support.

We evaluated the impact of a scaled-up CHW program on child survival in Uganda using elements of an adaptive design. A previous pilot version of the program, in which financially incentivized CHWs delivered integrated community case management (iCCM), maternal, newborn, and child health (MNCH) services, as well as basic treatment and preventive care, led to significant reductions in under-five and infant mortality (27% and 33%, respectively) after three years.[7]

Methods

Context

In 2007, two NGOs in Uganda piloted a novel community health initiative emphasizing data-driven performance management, routine in-service training, and supportive supervision to empower CHWs in delivering high-quality primary healthcare. Unlike many volunteer-based programs, this initiative introduced financial incentives for CHWs, employing a franchised direct-selling model to expand access to affordable, high-impact health products and essential maternal and child health services. Performance-based incentives further motivated key activities, including household visits, child assessments, maternal enrollment, and newborn

care. The program was structured into geographically defined branches, each affiliated with a specific NGO.

In 2010, the joint pilot program, operating in a few hundred villages, was evaluated through a cluster-randomized controlled trial. The evaluation documented a 27% reduction in under-five mortality in the treatment arm compared to the control arm over three years. Encouraged by these results, the two NGOs, in partnership with the Ministry of Health, expanded the program nationwide and embedded a new impact evaluation into the scale-up process.

While the pilot and scaled-up program designs shared many similarities, key differences existed. Monitoring intensity and CHW earnings were lower in the scaled-up program compared to the pilot. Additionally, the broader context had changed; for example, infant mortality rates across Uganda had declined by 24% at the start of the scale-up evaluation compared to the beginning of the pilot trial.

Study design and participants

The study was a parallel-group, stratified cluster-randomized controlled trial embedded within the nationwide expansion of the CHW program. A total of 500 rural and semi-rural clusters across 15 geographical zones in 13 districts, representing all four regions of Uganda, participated (Figure A1). Clusters were evenly split between two NGOs: 250 clusters across four zones for NGO A and 250 clusters across 11 zones for NGO B. Within each zone, clusters were randomly assigned to either the intervention or control group through a randomization procedure (computerized random numbers) generated by the researchers using Stata 12 (StataCorp, College Station, TX) statistical software, using balanced 1:1 randomization (Figure 1). Intervention clusters received at least one program CHW, while control clusters had no assigned program CHWs. Although CHWs were recommended to focus on households within their cluster, they were not restricted from serving households outside their assigned area.

Households in both intervention and control clusters could access primary healthcare from other sources, including private clinics and government health workers. By September 2016, the program was fully operational in all intervention clusters. At baseline, each cluster had an average of 182 households.

The primary aim of the study was to assess the impact of having at least one program CHW per cluster on all-cause under-five mortality. The evaluation was conducted independently of program implementation. Before the evaluation, all clusters were enumerated, and a baseline survey was conducted in a representative sample of approximately 25 eligible households per cluster. Eligibility was determined by the likelihood of having at least one child born during the study period. Pre-tested eligibility criteria, developed through local consultations, included households with a female resident aged 16–35 who was pregnant, had a young child, or was married. The female head of the household was the primary respondent for the baseline survey, conducted from January to June 2016.

Study outcomes were measured through a household survey conducted between March and June 2021, approximately five years after CHWs began operating in intervention clusters. Of the households identified at baseline, 80.6% were successfully tracked and surveyed, including those who remained in the same location (87.6%) and those who had migrated but were reachable (12.4%). The target respondents were the original baseline respondents or, if unavailable, the next most knowledgeable woman in the household. Overall, 83.3% of surveys were completed by the original respondents. Figure 1 summarizes the study design.

All surveys were conducted by Innovations for Poverty Action (IPA) Uganda using trained local staff with prior data collection experience. Data collectors were blinded to whether they were interviewing in intervention or control clusters. Surveys were administered digitally using SurveyCTO-Open Data Kit (ODK), with encrypted responses stored daily on a secure

server. Random back-checks ensured compliance with data collection protocols, and no violations were identified.

Trial design

Concerns about whether the significant effects observed during the pilot phase could be sustained at scale led to the incorporation of adaptive design elements in the scale-up trial. Specifically, a midline performance review was conducted approximately one year into the trial. This assessment focused on secondary study outcomes that could change relatively quickly and was measured through a household survey administered to all study households. The results were then shared with the two organizations responsible for program implementation.

While ongoing monitoring—such as through a Data and Safety Monitoring Board (DSMB)—is required in many clinical trials, midline reviews conducted by evaluators and shared with program implementers are uncommon in randomized field trials. The adaptive design approach offers potential benefits by enabling evidence-based adjustments to the intervention, implementation guidelines, and trial design while keeping the original treatment assignment unchanged. In essence, this design transforms the trial from a one-time evaluation into an iterative learning and experimentation process.

The performance review report, which indicated lower-than-expected improvements in secondary outcomes, was shared with the two NGOs in October 2018 and discussed in meetings in October and November 2018. In response, NGO A introduced major modifications to the design and implementation of the program (henceforth Program A), whereas NGO B introduced only minor adjustments to it (henceforth Program B).

In November 2018, the evaluators proposed extending the study by two years to allow sufficient time for the revised programs to affect outcomes. This proposal was approved by implementers and funders. The extension divided the study period into two phases (Figure A2).

Phase I (original program): from program launch across all intervention clusters (September 2016) to the start of midline data collection (October 2017). Phase II (revised program): from when NGOs started implementing program revisions (December 2018) to the start of endline data collection (March 2021), covering 28 months. While some program revisions took time to become fully operational, the dissemination of performance review findings allowed NGOs to react immediately. As shown in the results section, findings remain robust to minor adjustments in the exact definition of Phase II's starting point.

The trial was fully integrated into the CHW program scale-up. CHWs were blinded to a village's inclusion in the trial sample.

The trial was approved by the ethic committee of the Uganda National Council for Science and Technology (Reference No. SS3938), the IPA Institutional Review Board (Reference No. 13774) based in the USA, and the MILDMAY Institutional Review Board based in Uganda (Reference No. 0109-2015). The trial was registered in the Pan African Clinical Trials Registry (PACTR201609001398349) and in the American Economic Association RCT registry (AEARCTR-0002392) and the registrations can be accessed online (http://www.pactr.org and www.socialscienceregistry.org).

Outcomes

The primary outcome in the initial study design was all-cause under-five mortality measured over a three-year period. However, due to protocol and program revisions, we focus on the impact of the revised CHW programs.

As illustrated in Figure A3, child mortality during the 28-month Phase II period can be defined in multiple ways, depending on sample size, age composition, and average exposure to the revised program. One approach ("*All children under 5*") includes all children born up to 58 months before Phase II began, who were therefore potentially exposed to the revised program for at least one month before turning five. This approach maximizes sample size and

enables under-five mortality estimation but includes a large proportion of older children with lower mortality risk. Additionally, many of these children, even in treatment villages, were primarily exposed to the initial CHW program rather than the revised version.

An alternative approach ("*Born in Phase II*") includes only children born during Phase II. This sample consists exclusively of younger children (no older than 28 months) who, if in treatment villages, were exposed to the revised program since birth. Our primary mortality measure balances these two approaches, including children born during Phase II as well as those born up to 27 months before Phase II began ("*Under 28 months*"). This selection leads us to estimate mortality under 28 months of age over the Phase II period.

All mortality rates were calculated using household survey data collected at the trial's end. Depending on the study period and mortality definition, rates were computed as the number of deaths within a specific age range and time period, per 1,000 child-years of exposure to the risk of death.

Secondary outcomes included CHW interactions (program coverage), under-five morbidity from malaria, pneumonia, and diarrhea, treatment services for these illnesses, reproductive, antenatal, delivery, and postnatal care, basic health knowledge, and adoption of health prevention measures. Data for all secondary outcomes were collected through the endline household survey.

Statistical analysis

Intention-to-treat (ITT) analyses were conducted to compare intervention and control clusters for each outcome. ITT status was determined by the cluster of residence at baseline, as recorded in the baseline survey for successfully tracked households.

For mortality rates, we report rate ratios estimated using a Poisson model, adjusting for the stratified randomized design with binary zone indicators. Rate ratios were derived using the marginal standardization technique, with 95% confidence intervals (CIs) estimated via the

delta method. For all other outcomes, we report risk ratios adjusted for the stratified randomized design. Standard errors were clustered at the cluster level to account for intra-cluster correlation among households within the same cluster.

Statistical analyses were conducted using Stata 17 (StataCorp, College Station, Texas). Appendix B details sample size determination and power calculations.

Role of the funding source

The funding source was not involved in the design or the conduct of the trial, including data gathering, data analysis, and report writing.

Results

This section is divided into three parts. The first provides an overview of the scale-up operations of both NGOs and key statistics. The second highlights core findings from the midline performance review and the resulting program design changes implemented by the two NGOs. The third presents endline results, focusing on the impacts of the revised programs.

Scale up and summary statistics

Following the evaluation of their pilot program, both NGOs significantly expanded their operations in scale and scope (see Figure A4 for details on the operational scale-up over time). By 2016, at the start of the scale-up study, NGO A had more than tripled its branch offices across Uganda and increased its CHW workforce to over 1,600—an eightfold increase since 2011. Expansion continued, and by 2021, NGO A had approximately 4,500 active CHWs in the field. Similarly, by 2016, NGO B oversaw more than 4,000 CHWs across 134 branches, up from approximately 2,000 in 2011.

The rapid expansion in CHW numbers and geographical coverage posed challenges for supervision, monitoring, and training. Anticipating some of these difficulties, the NGOs introduced digital tools to improve reporting and oversight. NGO A implemented these tools

at the outset of the scale-up study, whereas NGO B faced delays and did not have the technology fully operational during the first study phase. Compared to the 2011–2013 proof-of-concept study, both NGOs reduced their emphasis on the program's revenue-generating aspects while increasing CHW compensation through pay-for-performance incentives. As a result, average CHW earnings declined. By 2016, CHWs earned approximately 45% less than at the start of the pilot program in 2011.

Table 1 presents balance tests by implementing organization using baseline data. Pre-trial household and child characteristics were not statistically different between the treatment and control groups. However, differences existed across the locations served by the two NGOs. Households in NGO B areas were generally larger, with lower levels of education and wealth than those in NGO A areas. Additionally, households in NGO B areas reported more healthcare provider visits in the previous 30 days compared to those in NGO A areas.

Midline: results and program changes

The trial design incorporated a midline performance review approximately one year into the three-year trial period, allowing corrective actions if process outcomes fell below expectations. The key performance indicator was the frequency of household interactions with CHWs.

As shown in Table 2, treatment households in both samples were more likely to have interacted with CHWs in the past 30 days compared to control households. However, the observed effect—a 5 percentage point increase (26% vs. 21% in the NGO A trial; 30% vs. 25% in the NGO B trial)—was significantly lower than expected. Notably, CHWs employed by the two NGOs played a limited role, accounting for only about a quarter of total CHW interactions in treatment groups (32% in NGO A and 20% in NGO B).

Consistent with these findings, there were no significant differences between treatment and control groups in antenatal or postnatal care, nor in the treatment of sick children, with two exceptions. In NGO A villages, newborns were 16 percentage points more likely to receive a

CHW visit within the first week (adjusted RR 2.83, 95% CI: 2.21–3.63). Additionally, children with diarrhea were 7 percentage points more likely to be treated with ORS/Zinc in NGO A villages (adjusted RR 1.19, 95% CI: 1.05–1.35) and 5 percentage points more likely to be treated with ORS/Zinc in NGO B villages (adjusted RR 1.06, 95% CI: 1.00–1.13).

Following the trial protocol, midline performance review findings were shared with the NGOs in late 2018. In response, both NGOs implemented several program modifications, though the scale and nature of the reforms varied significantly (see Table A1). To increase coverage, NGO A expanded its active CHW workforce by over 60%, while NGO B increased its workforce by 25% (Figure A5). Additionally, NGO A reduced the sale price of CHW-supplied medicines for malaria, diarrhea, and pneumonia, whereas no pricing changes were made in NGO B locations. Another key difference was in CHW compensation. NGO A raised monthly pay from $5–10 to $16–20 by expanding the pay-for-performance component, making sales margins less critical. In contrast, NGO B made only minor financial adjustments, keeping CHW compensation at approximately $8 per month. Supervision and monitoring approaches also diverged. NGO A improved oversight through a health app that enabled real-time, data-driven monitoring and support, while NGO B did not implement comparable digital oversight. These differences in adaptation resulted in distinct intervention models, despite the programs starting with similar designs.

Endline results

At endline, 81% of the original households in the NGO A trial sample and 80% in the NGO B trial sample were successfully tracked and surveyed. Attrition rates were comparable between treatment and control groups, with no statistically significant differences in the likelihood of attrition across assignment groups (see Table A2).

Table 3 presents the adjusted rate ratios describing the program's impact on child mortality during Phase II, stratified by NGO. Depending on the sample used, the revised CHW

program implemented by NGO A resulted in a 20–30% reduction in child mortality. Using our primary mortality measure, which includes all children born during Phase II as well as those born up to 27 months before Phase II began, child mortality fell by 28% (adjusted RR 0.72, 95% CI: 0.54–0.97). In contrast, no significant differences in child mortality were observed between intervention and control clusters in the NGO B trial.

Figure A6 in the Appendix visually illustrates the robustness of the mortality findings for NGO A by plotting child mortality estimates from a sample of children born during Phase II, progressively extending to include children born t = 1, 2,…58 months before Phase II began. Figure A7 further demonstrates that results remain robust regardless of the chosen starting month for Phase II. Notably, the estimated mortality reduction gets larger as one considers later starting months of Phase II, reflecting the time required for programmatic changes to take full effect in the field.

Table A3 presents results for the full five-year program, showing no significant differences in child mortality between intervention and control clusters in either NGO A or NGO B trials.

The CHW program aimed to provide households with young children with a broad range of child-specific health services and products. CHWs conducted home visits, educated families on essential health behaviors, provided basic medical advice, referred severe cases to health centers, visited newborns, and distributed or sold health products. Some activities, such as newborn visits within the first week and referrals of sick children to health clinics, were explicitly incentivized.

We measured both the quantity and, where possible, the quality of service provision in three key areas: (1) treatment of children for malaria, diarrhea, and pneumonia, (2) reproductive, antenatal, and postnatal care, and (3) health knowledge and prevention.

We begin by examining the first stage of the CHW program's causal pathway: the extent to which households regularly interacted with CHWs. Since no monitoring system was implemented in the study villages, we rely on data from the household survey conducted at the end of the trial.

Table 4 shows that in villages where NGO A was operating, 41% of treatment households had been visited by a CHW in the past 30 days. The majority of these visits (89%, or 933/1053) were conducted by the NGO's CHWs. This represents a significant increase compared to control households (adjusted RR 4.28, 95% CI: 3.59–5.11) and a substantial improvement from midline results. Increased CHW interactions were also observed in NGO B villages relative to control villages (adjusted RR 1.44, 95% CI: 1.25–1.66), although CHW visit rates were less than half of those in NGO A villages, and the proportion of visits conducted by NGO B's CHWs was less than a third.

Table 4 also reports the program's impact on health services for children who fell ill with malaria, diarrhea, or pneumonia. Self-reported morbidity rates were similar across treatment and control groups, but as shown in Figure A8, significant differences were observed in service provision, particularly in NGO A villages. Households exposed to Program A were six times more likely to have received a visit from a healthcare provider in the past three months (43% vs. 7%; adjusted RR 6.13, 95% CI: 5.02–7.48). They were also 8% more likely to have received correct treatment—ACT for malaria, ORS/Zinc for diarrhea, and antibiotics for pneumonia (87% vs. 81%; adjusted RR 1.08, 95% CI: 1.05–1.10). Additionally, they were eight times more likely to have received a follow-up check by a healthcare provider (37% vs. 5%; adjusted RR 7.82, 95% CI: 6.18–9.90) and 58% more likely to have received a referral to a health facility (5% vs. 3%; adjusted RR 1.58, 95% CI: 1.19–2.11). These improvements were observed across all three diseases.

For Program B, increases in service provision were smaller but still notable. Households in treatment villages were 28% more likely to have received a visit from a healthcare provider when a child was ill (23% vs. 18%; adjusted RR 1.28, 95% CI: 1.10–1.47). The likelihood of receiving correct treatment increased by 4% (90% vs. 87%; adjusted RR 1.04, 95% CI: 1.02–1.06), while follow-up visits increased by 77% (17% vs. 10%; adjusted RR 1.77, 95% CI: 1.46–2.14). However, the likelihood of receiving a referral (10% vs. 9%; adjusted RR 1.16, 95% CI: 0.96–1.40) did not differ significantly between treatment and control groups.

Table 4 also presents results on reproductive, antenatal, and postnatal care. In the NGO A trial, women in treatment villages were 8% more likely to have received family planning advice (67% vs. 62%; adjusted RR 1.08, 95% CI: 1.04–1.13), though no difference was found in actual family planning use (87% vs. 86%; adjusted RR 1.01, 95% CI: 0.98–1.03). Women in the treatment group were also 51% more likely to have been advised on the importance of antenatal care (20% vs. 13%; adjusted RR 1.51, 95% CI: 1.28–1.79), though no difference was found in the likelihood to attend at least four antenatal visits (42% vs. 38%; adjusted RR 1.09, 95% CI: 0.99–1.19).

During pregnancy, women in NGO A treatment villages were six times more likely to have received a visit from a health worker (39% vs. 6%; adjusted RR 6.53, 95% CI: 4.92–8.66). No significant differences were observed in health facility deliveries or use of safe delivery kits. However, women in treatment villages were five times more likely to have received a postnatal visit within the first week (33% vs. 6%; adjusted RR 5.53, 95% CI: 4.12–7.43) and 22% more likely to have received child nutrition advice (56% vs. 46%; adjusted RR 1.22, 95% CI: 1.14–1.30).

Improvements were also observed in Program B villages, though the effects were smaller. Women in treatment villages were twice as likely to have received a visit from a health

worker during pregnancy (20% vs. 9%; adjusted RR 2.22, 95% CI: 1.80–2.74) and 50% more likely to have received a postnatal visit (14% vs. 10%; adjusted RR 1.47, 95% CI: 1.19–1.81). However, these increases were less pronounced than those in the NGO A trial (see Table 4).

Finally, Table 4 shows that the revised CHW programs improved health knowledge and preventive behaviors in both treatment groups, though effect sizes were slightly larger for households exposed to Program A. Overall, these effects were modest, likely because most households already possessed basic health knowledge and adhered to standard preventive practices.

No harm or unintended effects of the intervention were reported.

Discussion

Maintaining the effectiveness of successful proof-of-concept programs during scale-up remains a key policy challenge. [8-10] In this paper, we present evidence from the scale-up of a CHW program in Uganda, using an adaptive design that allowed for ongoing adjustments throughout the evaluation. We assessed the impact of the scaled-up program on all-cause child mortality while keeping the same context and set of implementers as in the pilot study. Our findings provide valuable insights into how program adaptation can help sustain effectiveness during scale-up and contribute to the global discourse on expanding successful pilot initiatives.

Our results suggest that a flexible, adaptive approach during scale-up, combined with implementers willing to make evidence-based adjustments, can help sustain the success of pilot programs. Data-driven modifications enable implementers to continuously learn, refine the program, and increase the likelihood of successful scale-up. Moreover, incorporating evaluation metrics that predict key performance outcomes enhances the ability to monitor progress and implement timely improvements.

This study had several limitations. First, the decision not to implement a surveillance or monitoring system in the study villages preserved external validity but required relying on

retrospective recall data. The data collection protocol followed standardized methodologies, and any potential recall bias was expected to affect intervention and control groups equally, leading to attenuation bias rather than systematic biases.

Second, contamination between clusters due to geographic proximity was possible, likely leading to an underestimation of the program's true effects. Third, the trial was not originally designed to compare the effectiveness of the two implementers, and the midline revisions affected the study's statistical power to assess the program's impact.

Fourth, the two NGOs operated in slightly different contexts, which may have influenced the effectiveness of each CHW program and the impact of midline program adjustments. While the midline results fell below expectations, some indicators showed signs of improvement in intervention villages (e.g., increased postnatal visits in NGO A areas and improved diarrhea treatment in NGO B areas). Phase II results should therefore be interpreted in the context of these improvements.

Finally, the research team did not have access to the internal decision-making processes of the two NGOs, which resulted in NGO A making significant program modifications after identifying performance gaps, while NGO B implemented more modest changes. These difference choices may be explained by organizational priorities: NGO A specialized in community health services, whereas NGO B managed a broader portfolio of poverty-reduction and development programs. Given the midline findings, reallocating resources from a lower performing program to more promising initiatives may have been a strategic decision for NGO B.

More generally, assessing impact during scale-up and sharing findings helps implementers make informed decisions, not only about improving programs but also about scaling down or reallocating resources to more effective initiatives.

Panel 1: Research in context

Systematic review

PubMed for trials studying the impact of community health workers in developing countries, with a focus on child mortality. The search terms included combinations of the keywords "child," "developing," "cluster," "trial," and "mortality," alongside common terms used to refer to providers of community-based health services ("community health worker," "community health promoter," "village health worker," and "lay health worker").

The initial search resulted in a total of 329 studies. After filtering out non-experimental studies, those not conducted in developing countries, and those not focusing on community-based health interventions, we were left with 105 studies. These articles were further filtered by outcomes, excluding studies that did not look at child mortality. A final set of 21 relevant studies was then analyzed systematically.

Evidence before this study

Most studies (18) focused on the impact of community-based intervention on neonatal mortality, showing promising results. Eight studies assessed the impact of maternal training programs delivered through frontline health workers in South Asian countries. Four studies looked at tailored intervention strategies, designed following the identification of local perinatal challenges through women's groups, and found reductions in neonatal mortality ranging from 7-32%. [11-14] The other four studies examined more structured, pre-designed training approaches. Three interventions emphasized maternal health education, clean delivery practices, complication recognition, and newborn care, and estimated a 25-50% reduction in neonatal mortality.[15-17] One study trained mothers of low birth weight infants in skin-to-skin contact and found no significant effects instead. [18]

Ten additional studies assessed the impact of home visits by trained community members on neonatal mortality. Out of these, five were proof-of-principle studies, with four reporting significant reductions in mortality ranging between 34-78%,[4,9, 19, 20] while one found no significant effects.[21] The remaining five studies were conducted in program settings, with three documenting smaller reductions in neonatal mortality (9-28%),[22-24] and two finding no impact.[25,26]

Evidence on the effectiveness of frontline health worker programs on children beyond infancy remains scarce, with the three studies identified in our literature review finding no significant reduction in mortality.[27-29]

The variation in the estimated effects, particularly for home visit interventions delivered in program settings, suggests that the impact of frontline health worker programs depends heavily on effective management and implementation strategies. Recognizing that pilot studies often overestimate program effectiveness—since small-scale interventions benefit from intensive training and supervision that may not be feasible at scale—most studies focused on evaluating mid-sized and large-scale programs. Among the 21 studies analyzed, 19 involved large study populations, ranging from 10,000 to 503,163 individuals.

One key challenge for bringing programs to scale is the need to integrate them within the existing health system, which is essential for ensuring long-term sustainability and cost-effectiveness. Yet, even among the large-scale studies reviewed, community health programs were not always implemented in partnership with public health systems—only 11 studies examined initiatives led by or closely coordinated with public health providers.

Given the many challenges associated with scaling up successful programs, it is important to understand whether and how effectiveness can be preserved. However, only three studies looked at interventions that were being expanded from an initial pilot.[11, 15, 18] Moreover, in two out of three cases the original implementation was not rigorously assessed, making it impossible to compare the effectiveness between the pilot and the scaled-up version. Only one study provides this type of comparison, finding that the reduction in stillbirths and neonatal mortality was lower in the scaled-up version compared to the initial pilot.[15] As a consequence, there is also limited evidence on strategies to facilitate a smooth transition to scale. Only one study made slight program adaptations based on pilot insights, but it did not incorporate any form of adaptation strategy that could allow for adjustments in the program delivery throughout the scaling-up phase.[15]

Added value of this study:
This study offers novel insights into the effectiveness of a large-scale CHW program implemented by NGOs in collaboration with the Ugandan government. It contributes to the literature on CHW models and child mortality in three key ways. First, it expands the evidence on the impact of these programs beyond the neonatal and infant periods by examining child mortality across different age groups. Second, it strengthens the still limited body of evidence on the scalability of such programs, highlighting their feasibility for nationwide implementation through partnerships between governmental and non-governmental actors. Third, it explores how evidence-based adaptation strategies can preserve the effectiveness of a program as it scales. To our knowledge, no previous studies on CHW programs have examined how mid-way evaluations during scale-up efforts can inform program expansion and help maintain the effectiveness observed in pilot phases. The study underscores the importance of a rigorously evaluated learning phase in real-world conditions and illustrates how structured adaptation can bridge the gap between pilot projects and full-scale national implementation

References

1. Singh P, Sachs JD. 1 million community health workers in sub-Saharan Africa by 2015. *Lancet* 2013, 382: 363-365.
2. Tulenko K, Møgedal S, Afzal M, Frymus D, Oshin A, Pate M, et al. Community health workers for universal health-care coverage: From fragmentation to synergy. Bull World Health Organ. 2013; 91:847–52.
3. WHO guideline on health policy and system support to optimize community health worker programmes. Geneva: World Health Organization; 2018.
4. Kumar V, Mohanty S, Kumar A, for the Saksham Study Group, et al. Effect of community-based behaviour change management on neonatal mortality in Shivgarh, Uttar Pradesh, India: a cluster-randomised controlled trial. *Lancet* 2008, 372, pp. 1151–1162.
5. Baqui AH, El-Arifeen S, Darmstadt GL, for the Projahnmo Study Group, et al. Effect of community-based newborn-care intervention package implemented through two service-delivery strategies in Sylhet district, Bangladesh: a cluster-randomised controlled trial. *Lancet* 2008, 371: 1936–1944.
6. Satav AR, Satav KA, Bharadwaj A, et al. Effect of home-based childcare on childhood mortality in rural Maharashtra, India: a cluster randomised controlled trial. BMJ Glob Health. 2022; 7(7): e008909.
7. Björkman Nyqvist M, Guariso A, Svensson J, and D Yanagizawa-Drott. 2019. Reducing Child Mortality in the Last Mile: Experimental Evidence on Community Health Promoters in Uganda. *American Economic Journal: Applied Economics* 11(3): 155-92.
8. Al-Ubaydli, O., List, J.A. and Suskind, D. (2020), 2017 Klein lecture: the science of using science: toward an understanding of the threats to scalability. *International Economic Review*, 61: 1387-1409.
9. List J A (2022), *The voltage effect*. Crown Currency Publishing.
10. List J A (2024), "Optimally generate policy-based evidence before scaling," *Nature* 626: 491–499.
11. Azad K, Barnett S, Banerjee B, Shaha S, Khan K, Rego AR, Barua S, Flatman D, Pagel C, Prost A, Ellis M, Costello A. Effect of scaling up women's groups on birth outcomes in three rural districts in Bangladesh: a cluster-randomised controlled trial. *Lancet*. 2010 Apr 3;375(9721):1193-202. doi: 10.1016/S0140-6736(10)60142-0. Epub 2010 Mar 6. PMID: 20207412.
12. Manandhar DS, Osrin D, Shrestha BP, Mesko N, Morrison J, Tumbahangphe KM, Tamang S, Thapa S, Shrestha D, Thapa B, Shrestha JR, Wade A, Borghi J, Standing H, Manandhar M, Costello AM; Members of the MIRA Makwanpur trial team. Effect of a participatory intervention with women's groups on birth outcomes in Nepal: cluster-randomised controlled trial. *Lancet*. 2004 Sep 11-17;364(9438):970-9. doi: 10.1016/S0140-6736(04)17021-9. PMID: 15364188.
13. Tripathy P, Nair N, Barnett S, Mahapatra R, Borghi J, Rath S, Rath S, Gope R, Mahto D, Sinha R, Lakshminarayana R, Patel V, Pagel C, Prost A, Costello A. Effect of a participatory intervention with women's groups on birth outcomes and maternal depression in Jharkhand and Orissa, India: a cluster-randomised controlled trial. *Lancet*. 2010 Apr 3;375(9721):1182-92. doi: 10.1016/S0140-6736(09)62042-0. Epub 2010 Mar 6. PMID: 20207411.
14. Tripathy P, Nair N, Sinha R, Rath S, Gope RK, Rath S, Roy SS, Bajpai A, Singh V, Nath V, Ali S, Kundu AK, Choudhury D, Ghosh SK, Kumar S, Mahapatra R, Costello A, Fottrell E, Houweling TA, Prost A. Effect of participatory women's groups facilitated by Accredited Social Health Activists on birth outcomes in rural eastern India: a cluster-randomised controlled trial. *Lancet Glob Health*. 2016 Feb;4(2):e119-28. doi: 10.1016/S2214-109X(15)00287-9. PMID: 26823213.
15. Bhutta ZA, Soofi S, Cousens S, Mohammad S, Memon ZA, Ali I, Feroze A, Raza F, Khan A, Wall S, Martines J. Improvement of perinatal and newborn care in rural Pakistan through community-based strategies: a cluster-randomised effectiveness trial. *Lancet*. 2011 Jan 29;377(9763):403-12. doi: 10.1016/S0140-6736(10)62274-X. Epub 2011 Jan 14. PMID: 21239052.
16. Gai Tobe R, Haque SE, Mubassara S, Rahman R, Ikegami K, Mori R. Maternal and child health handbook to improve continuum of maternal and child care in rural Bangladesh: Findings of a cluster randomized controlled trial. *PLoS One*. 2022 Apr 6;17(4):e0266074. doi: 10.1371/journal.pone.0266074. PMID: 35385542; PMCID: PMC8986009.
17. Midhet F, Becker S. Impact of community-based interventions on maternal and neonatal health indicators: Results from a community randomized trial in rural Balochistan, Pakistan. *Reprod Health*. 2010 Nov 5;7:30. doi: 10.1186/1742-4755-7-30. PMID: 21054870; PMCID: PMC2993657.
18. Sloan NL, Ahmed S, Mitra SN, Choudhury N, Chowdhury M, Rob U, Winikoff B. Community-based kangaroo mother care to prevent neonatal and infant mortality: a randomized, controlled cluster trial. *Pediatrics*. 2008 May;121(5):e1047-59. doi: 10.1542/peds.2007-0076. PMID: 18450847.
19. Baqui AH, Arifeen SE, Williams EK, Ahmed S, Mannan I, Rahman SM, Begum N, Seraji HR, Winch PJ, Santosham M, Black RE, Darmstadt GL. Effectiveness of home-based management of newborn

infections by community health workers in rural Bangladesh. *Pediatr Infect Dis J*. 2009 Apr;28(4):304-10. doi: 10.1097/INF.0b013e31819069e8. PMID: 19289979; PMCID: PMC2929171.
20. Lewycka S, Mwansambo C, Rosato M, Kazembe P, Phiri T, Mganga A, Chapota H, Malamba F, Kainja E, Newell ML, Greco G, Pulkki-Brännström AM, Skordis-Worrall J, Vergnano S, Osrin D, Costello A. Effect of women's groups and volunteer peer counselling on rates of mortality, morbidity, and health behaviours in mothers and children in rural Malawi (MaiMwana): a factorial, cluster-randomised controlled trial. *Lancet*. 2013 May 18;381(9879):1721-35. doi: 10.1016/S0140-6736(12)61959-X. PMID: 23683639; PMCID: PMC3796349.
21. Ciccone EJ, Hu D, Preisser JS, Cassidy CA, Kabugho L, Emmanuel B, Kibaba G, Mwebembezi F, Juliano JJ, Mulogo EM, Boyce RM. Point-of-care C-reactive protein measurement by community health workers safely reduces antimicrobial use among children with respiratory illness in rural Uganda: A stepped wedge cluster randomized trial. *PLoS Med*. 2024 Aug 19;21(8):e1004416. doi: 10.1371/journal.pmed.1004416. PMID: 39159269; PMCID: PMC11407643.
22. Bhandari N, Mazumder S, Taneja S, Sommerfelt H, Strand TA; IMNCI Evaluation Study Group. Effect of implementation of Integrated Management of Neonatal and Childhood Illness (IMNCI) programme on neonatal and infant mortality: cluster randomised controlled trial. *BMJ*. 2012 Mar 21;344:e1634. doi: 10.1136/bmj.e1634. PMID: 22438367; PMCID: PMC3309879.
23. Bhutta ZA, Memon ZA, Soofi S, Salat MS, Cousens S, Martines J. Implementing community-based perinatal care: results from a pilot study in rural Pakistan. *Bull World Health Organ*. 2008 Jun;86(6):452-9. doi: 10.2471/blt.07.045849. PMID: 18568274; PMCID: PMC2647462.
24. Soofi S, Ariff S, Sadiq K, Habib A, Bhatti Z, Ahmad I, Hussain M, Ali N, Cousens S, Bhutta ZA. Evaluation of the uptake and impact of neonatal vitamin A supplementation delivered through the Lady Health Worker programme on neonatal and infant morbidity and mortality in rural Pakistan: an effectiveness trial. *Arch Dis Child*. 2017 Mar;102(3):216-223. doi: 10.1136/archdischild-2016-310542. Epub 2016 Jun 28. PMID: 27471856.
25. Darmstadt GL, Choi Y, Arifeen SE, Bari S, Rahman SM, Mannan I, Seraji HR, Winch PJ, Saha SK, Ahmed AS, Ahmed S, Begum N, Lee AC, Black RE, Santosham M, Crook D, Baqui AH; Bangladesh Projahnmo-2 Mirzapur Study Group. Evaluation of a cluster-randomized controlled trial of a package of community-based maternal and newborn interventions in Mirzapur, Bangladesh. *PLoS One*. 2010 Mar 24;5(3):e9696. doi: 10.1371/journal.pone.0009696. PMID: 20352087; PMCID: PMC2844410.
26. Rasaily R, Saxena NC, Pandey S, Garg BS, Swain S, Iyengar SD, Das V, Sinha S, Gupta S, Sinha A, Kumar S, Pandey A, Pandey RM, Sachdev HS, Sankar MJ, Ramji S, Paul VK, Bang AT; ICMR-HBMYI Study Group; ICMR -HBMYI Study Group. Effect of home-based newborn care on neonatal and infant mortality: a cluster randomised trial in India. *BMJ Glob Health*. 2020 Sep;5(9):e000680. doi: 10.1136/bmjgh-2017-000680. PMID: 32972965; PMCID: PMC7517550.
27. Boone P, Elbourne D, Fazzio I, Fernandes S, Frost C, Jayanty C, King R, Mann V, Piaggio G, dos Santos A, Walker PR. Effects of community health interventions on under-5 mortality in rural Guinea-Bissau (EPICS): a cluster-randomised controlled trial. *Lancet Glob Health*. 2016 May;4(5):e328-35. doi: 10.1016/S2214-109X(16)30048-1. PMID: 27102196.
28. Chinbuah MA, Kager PA, Abbey M, Gyapong M, Awini E, Nonvignon J, Adjuik M, Aikins M, Pagnoni F, Gyapong JO. Impact of community management of fever (using antimalarials with or without antibiotics) on childhood mortality: a cluster-randomized controlled trial in Ghana. *Am J Trop Med Hyg*. 2012 Nov;87(5 Suppl):11-20. doi: 10.4269/ajtmh.2012.12-0078. PMID: 23136273; PMCID: PMC3748510.
29. Pence BW, Nyarko P, Phillips JF, Debpuur C. The effect of community nurses and health volunteers on child mortality: the Navrongo Community Health and Family Planning Project. *Scand J Public Health*. 2007;35(6):599-608. doi: 10.1080/14034940701349225. PMID: 17852975.

Figure 1. CONSORT flow diagram

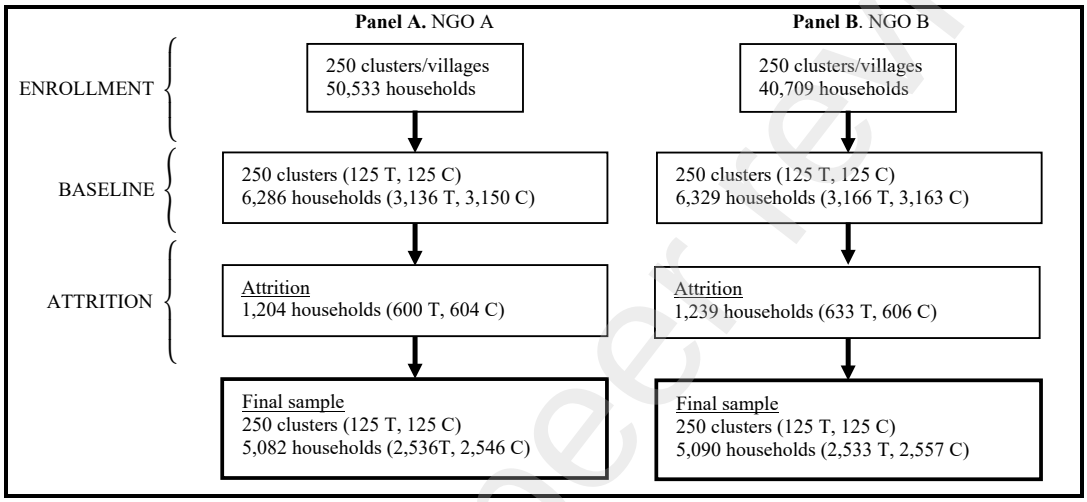


Table 1. Balance Checks

	NGO A				NGO B			
	Mean (SE)		Difference	Obs.	Mean (SE)		Difference	Obs.
Variable	Control	Treatment	(p-value)		Control	Treatment	(p-value)	
Age female HH head	23.410	23.549	0.138	6,286	23.319	23.244	-0.071	6,329
	(3.644)	(3.721)	[0.204]		(3.673)	(3.695)	[0.522]	
Female HH head can read	0.677	0.694	0.016	6,285	0.573	0.572	-0.001	6,327
	(0.468)	(0.461)	[0.373]		(0.495)	(0.495)	[0.958]	
# years living in the village	7.204	7.153	-0.052	6,274	6.528	6.477	-0.062	6,299
	(7.222)	(7.185)	[0.811]		(6.319)	(5.960)	[0.723]	
HH asset index (PCA)	0.386	0.477	0.091	6,074	-0.424	-0.451	-0.021	5,987
	(1.855)	(1.958)	[0.437]		(1.685)	(1.588)	[0.814]	
# adult (18+) HH members	2.327	2.315	-0.011	6,286	2.531	2.537	0.004	6,329
	(0.896)	(0.939)	[0.687]		(1.160)	(1.130)	[0.915]	
# HH members of age 5-17	1.589	1.635	0.047	6,286	2.474	2.475	-0.011	6,329
	(1.684)	(1.759)	[0.366]		(2.821)	(2.792)	[0.886]	
# children under-5	1.513	1.511	-0.001	6,286	1.394	1.425	0.031	6,329
	(0.803)	(0.839)	[0.951]		(0.803)	(0.790)	[0.160]	
Death child under 5 past year	0.027	0.026	-0.000	5,684	0.023	0.017	-0.006	5,588
	(0.161)	(0.160)	[0.931]		(0.150)	(0.128)	[0.080]	
Currently pregnant	0.218	0.209	-0.010	6,286	0.186	0.184	-0.001	6,329
	(0.413)	(0.407)	[0.390]		(0.389)	(0.388)	[0.894]	
Ever used family planning	0.650	0.666	0.015	6,285	0.543	0.553	0.009	6,326
	(0.477)	(0.472)	[0.329]		(0.498)	(0.497)	[0.528]	
Currently uses family planning	0.374	0.384	0.010	5,685	0.293	0.284	-0.010	5,949
	(0.484)	(0.486)	[0.509]		(0.455)	(0.451)	[0.433]	
Discuss plans for child w/ husband	0.558	0.537	-0.021	6,252	0.524	0.542	0.018	6,310
	(0.497)	(0.499)	[0.106]		(0.499)	(0.498)	[0.204]	
Health care provider visit past 30d	0.069	0.059	-0.009	6,276	0.289	0.289	0.002	6,287
	(0.253)	(0.236)	[0.215]		(0.453)	(0.453)	[0.906]	
VHT visited HH past 30d	0.028	0.026	-0.002	6,276	0.151	0.161	0.011	6,287
	(0.164)	(0.158)	[0.670]		(0.358)	(0.367)	[0.360]	
# health service centers within 1h	2.071	2.181	0.111	6,286	1.883	1.924	0.041	6,329
	(1.266)	(1.301)	[0.153]		(1.109)	(1.169)	[0.436]	
Any child born last year	0.381	0.375	-0.006	6,224	0.343	0.351	0.008	6,305
	(0.486)	(0.484)	[0.666]		(0.475)	(0.477)	[0.532]	
Visited by health worker for ANC	0.023	0.029	0.005	1,347	0.049	0.041	-0.008	1,186
	(0.151)	(0.168)	[0.578]		(0.216)	(0.198)	[0.541]	
Delivery w/ medical help recommended	0.631	0.638	0.006	5,442	0.769	0.769	-0.001	5,381
	(0.483)	(0.481)	[0.698]		(0.421)	(0.422)	[0.946]	
Antimalarial treatment durign pregnancy	0.823	0.823	-0.001	5,394	0.782	0.793	0.008	5,339
	(0.382)	(0.382)	[0.955]		(0.413)	(0.405)	[0.497]	
Baby was exclsuively breastfed	0.436	0.428	-0.009	3,554	0.490	0.480	-0.008	3,609
	(0.496)	(0.495)	[0.619]		(0.500)	(0.500)	[0.675]	
Visited by health worker in first week	0.071	0.086	0.014	5,369	0.081	0.075	-0.006	5,192
	(0.257)	(0.280)	[0.188]		(0.273)	(0.264)	[0.528]	
Age in months	26.090	26.048	-0.042	9,185	26.271	25.962	-0.320	8,618
	(16.793)	(16.579)	[0.893]		(16.722)	(16.505)	[0.322]	
Girl	0.489	0.489	-0.000	9,185	0.483	0.495	0.012	8,618
	(0.500)	(0.500)	[0.977]		(0.500)	(0.500)	[0.274]	
Height-for-age (z-score)	-0.662	-0.689	-0.029	9,185	-0.369	-0.294	0.076	8,618
	(1.505)	(1.503)	[0.589]		(1.489)	(1.479)	[0.116]	
Weight-for-height (z-score)	-0.444	-0.465	-0.019	8,980	-0.673	-0.698	-0.026	8,318
	(1.139)	(1.120)	[0.573]		(1.186)	(1.146)	[0.483]	
MUAC	14.682	14.713	0.032	9,184	14.713	14.767	0.054	8,612
	(1.414)	(1.450)	[0.462]		(1.436)	(1.435)	[0.205]	
Child has insecticide treated net	0.622	0.628	0.006	9,297	0.781	0.783	0.002	8,800
	(0.485)	(0.483)	[0.718]		(0.413)	(0.412)	[0.864]	
Health care provider visit for Malaria	0.033	0.028	-0.005	5,263	0.050	0.055	0.006	5,512
	(0.179)	(0.165)	[0.367]		(0.218)	(0.228)	[0.480]	
Received referral for Malaria	0.172	0.179	0.006	5,453	0.091	0.088	-0.002	5,718
	(0.377)	(0.383)	[0.714]		(0.288)	(0.283)	[0.837]	
Health care provider visit for Pneumonia	0.036	0.024	-0.011	3,141	0.029	0.038	0.010	3,122
	(0.186)	(0.152)	[0.114]		(0.167)	(0.191)	[0.202]	
Received referral for Pneumonia	0.174	0.186	0.012	3,485	0.062	0.055	-0.007	3,609
	(0.379)	(0.390)	[0.537]		(0.242)	(0.228)	[0.525]	
Health care provider visit for Diarrhea	0.027	0.031	0.004	2,355	0.030	0.052	0.02	2,705
	(0.161)	(0.173)	[0.593]		(0.172)	(0.223)	[0.062]	
Received referral for Diarrhea	0.124	0.138	0.012	3,028	0.096	0.076	-0.02	3,150
	(0.329)	(0.345)	[0.478]		(0.294)	(0.264)	[0.092]	

Notes: The table shows the mean and standard error of the mean (in parenthesis) for different variables. The sample consists of all households surveyed at baseline in areas covered by NGO A (N=6,286) and by NGO B (N=6,329). Columns 3 and 7 show the difference in means and the p-value of the test for means equality (H0:=mean is equal across treatment and control). The test for equality takes into account district-fixed effects and clustering at the cluster (village) level.

Table 2. Process indicator outcomes at Midline

	NGO A			NGO B		
	Intervention	**Control**	**Adj RR (95% CI)**	**Intervention**	**Control**	**Adj RR (95% CI)**
Household interaction with CHWs						
Visited by a CHW (last 30 days)	756/2855 (26%)	585/2833 (21%)	1.30 (1.13 - 1.50)	876/2892 (30%)	702/2848 (25%)	1.23 (1.08 - 1.39)
Visited by an NGO CHW (last 30 days)	238/2855 (8%)	19/2833 (1%)	12.63 (6.60 - 24.20)	171/2892 (6%)	6/2848 (0%)	28.04 (12.51 - 62.85)
Antenatal and Postnatal care						
Received any antenatal care	1709/1925 (89%)	1730/1940 (89%)	1.00 (0.97 - 1.02)	1805/2009 (90%)	1769/1942 (91%)	0.99 (0.97 - 1.01)
...At least 4 antenatal care visits	1105/1706 (65%)	1110/1728 (64%)	1.01 (0.95 - 1.07)	1264/1804 (70%)	1223/1768 (69%)	1.01 (0.96 - 1.06)
Received postnatal visit in first week	366/1451 (25%)	135/1497 (9%)	2.83 (2.21 - 3.63)	336/1545 (22%)	274/1533 (18%)	1.18 (0.96 - 1.45)
Morbidity and treatment of sick children						
Child had malaria (last 3 months)	2742/4301 (64%)	2672/4418 (60%)	1.05 (1.00 - 1.10)	2798/4379 (64%)	2734/4268 (64%)	1.00 (0.96 - 1.04)
...Treated with ACT	2160/2662 (81%)	2097/2589 (81%)	1.00 (0.97 - 1.04)	2430/2743 (89%)	2358/2696 (87%)	1.01 (0.99 - 1.04)
Child had diarrhea (last 3 months)	1590/4301 (37%)	1629/4421 (37%)	1.00 (0.94 - 1.06)	1824/4376 (42%)	1890/4268 (44%)	0.94 (0.88 - 1.01)
...Treated with ORS/Zinc	518/1168 (44%)	435/1176 (37%)	1.19 (1.05 - 1.35)	1099/1578 (70%)	1056/1622 (65%)	1.06 (1.00 - 1.13)
Child had pneumonia (last 3 months)	2143/4302 (50%)	2133/4421 (48%)	1.03 (0.96 - 1.11)	2363/4381 (54%)	2387/4269 (56%)	0.96 (0.90 - 1.02)
...Treated with antibiotic	796/1939 (41%)	742/1890 (39%)	1.05 (0.92 - 1.19)	1308/2098 (62%)	1375/2080 (66%)	0.94 (0.88 - 1.00)

Notes: Data are n (%) from midline sample household survey. Group assignment is based on the baseline cluster of residence of the household. Shares are computed relative to the total valid answers (missing answers are excluded). Antenatal and Postnatal care refer to pregnancies and deliveries since the start of the program. Adjusted risk ratios are computed using a Poisson model, adjusting for stratified randomization. Confidence intervals are constructed using robust standard errors clustered at the cluster (village) level.

Table 3. Program impact on child mortality

	NGO A			NGO B		
	Intervention	**Control**	**Adj RR (95% CI)**	**Intervention**	**Control**	**Adj RR (95% CI)**
Panel A: Under-28 months mortality for those exposed to the risk of death during the exposure period						
Years of exposure to the risk of death	3882	3953		3824	3888	
# child deaths	84	119		77	74	
Mortality rate per 1,000 years of exposure	21.85	30.22	0.72 (0.54 - 0.97)	20.63	18.65	1.05 (0.78 - 1.42)
Panel B: Under-5 years mortality for those exposed to the risk of death during the exposure period						
Years of exposure to the risk of death	9374	9600		9035	9186	
# child deaths	109	140		89	92	
Mortality rate per 1,000 years of exposure	11.56	14.53	0.80 (0.62 - 1.03)	9.94	9.55	0.98 (0.73 - 1.30)
Panel C: All children born and exposed to the risk of death during the exposure period						
Years of exposure to the risk of death	1872	1886		1850	1906	
# child deaths	64	93		71	62	
Mortality rate per 1,000 years of exposure	35.95	52.56	0.70 (0.50 - 0.97)	44.26	33.32	1.19 (0.85 - 1.67)

Notes: This Table replicates Table 3 in the main text, excluding replacement households. Group assignment is based on the baseline cluster of residence of the household. The table report mortality during Phase II, which runs from December 2021 until March 2021. Adjusted risk ratios are computed using a Poisson model, adjusting for stratified randomization. Confidence intervals are constructed using robust standard errors clustered at the cluster (village) level. As a robustness check, Replacement households were excluded from the sample.

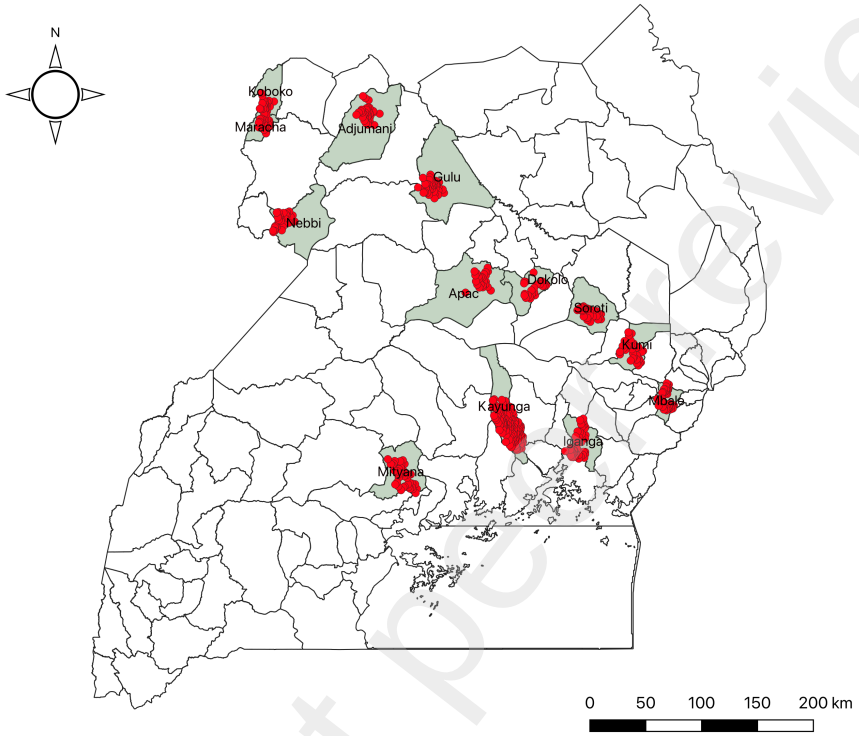
Table 4. Process indicator outcomes at Endline

	NGO A			NGO B		
	Intervention	**Control**	**Adj RR (95% CI)**	**Intervention**	**Control**	**Adj RR (95% CI)**
Household interaction with CHWs						
Visited by a CHW in the last 30 days	1053/2562 (41%)	246/2560 (10%)	4.28 (3.59 - 5.11)	528/2541 (21%)	369/2563 (14%)	1.44 (1.25 - 1.66)
Visited by an NGO CHW in the last 30 days	933/2562 (36%)	50/2560 (2%)	18.65 (13.67 - 25.44)	254/2541 (10%)	29/2563 (1%)	8.86 (5.87 - 13.38)
Morbidity and treatment of sick children						
Child had malaria, last 3 months	1941/3127 (62%)	1805/3156 (57%)	1.09 (1.03 - 1.15)	2223/3073 (72%)	2253/3122 (72%)	1.00 (0.97 - 1.03)
Child had diarrhea, last 3 months	998/3120 (32%)	994/3153 (32%)	1.02 (0.94 - 1.10)	1017/3054 (33%)	972/3103 (31%)	1.06 (0.99 - 1.14)
Child had pneumonia, last 3 months	1230/3125 (39%)	1288/3159 (41%)	0.96 (0.89 - 1.05)	1257/3062 (41%)	1329/3108 (43%)	0.95 (0.88 - 1.02)
Child had one of the above, last 3 months	2526/3132 (81%)	2518/3163 (80%)	1.01 (0.98 - 1.04)	2617/3077 (85%)	2632/3126 (84%)	1.01 (0.99 - 1.03)
...received medication (at least once)	2206/2526 (87%)	2045/2518 (81%)	1.08 (1.05 - 1.10)	2366/2617 (90%)	2289/2632 (87%)	1.04 (1.02 - 1.06)
...visited by a healthcare provider (at least once)	1076/2526 (43%)	176/2518 (7%)	6.13 (5.02 - 7.48)	612/2617 (23%)	480/2632 (18%)	1.28 (1.10 - 1.47)
...follow-up by healthcare provider (at least once)	930/2526 (37%)	119/2518 (5%)	7.82 (6.18 - 9.90)	453/2617 (17%)	256/2632 (10%)	1.77 (1.46 - 2.14)
...referral to health facility (at least once)	131/2526 (5%)	83/2518 (3%)	1.58 (1.19 - 2.11)	273/2617 (10%)	236/2632 (9%)	1.16 (0.96 - 1.40)
Reproductive, Antenatal, and Postnatal care						
Received advice on family planning practices	1724/2562 (67%)	1590/2560 (62%)	1.08 (1.04 - 1.13)	1931/2541 (76%)	1954/2563 (76%)	1.00 (0.96 - 1.03)
Ever used family planning	2229/2562 (87%)	2210/2560 (86%)	1.01 (0.98 - 1.03)	1982/2541 (78%)	2010/2563 (78%)	0.99 (0.96 - 1.03)
Advised to seek antenatal care	331/1636 (20%)	222/1663 (13%)	1.51 (1.28 - 1.79)	406/1814 (22%)	421/1894 (22%)	1.00 (0.87 - 1.16)
Sought antenatal care	1457/1646 (89%)	1499/1666 (90%)	0.98 (0.96 - 1.01)	1697/1825 (93%)	1776/1902 (93%)	1.00 (0.98 - 1.01)
At least 4 antenatal care visits	688/1646 (42%)	639/1666 (38%)	1.09 (0.99 - 1.19)	948/1825 (52%)	971/1902 (51%)	1.02 (0.95 - 1.09)
Visited by health worker during pregnancy	637/1641 (39%)	99/1668 (6%)	6.53 (4.92 - 8.66)	358/1808 (20%)	166/1888 (9%)	2.22 (1.80 - 2.74)
Advised to give birth with qualified help	177/265 (67%)	146/255 (57%)	1.16 (1.00 - 1.34)	119/153 (78%)	86/137 (63%)	1.26 (1.08 - 1.48)
Delivered in facility or with safe delivery kit	974/1320 (74%)	972/1343 (72%)	1.02 (0.95 - 1.09)	1325/1520 (87%)	1362/1583 (86%)	1.01 (0.98 - 1.05)
Received postnatal visit within first week	440/1316 (33%)	82/1341 (6%)	5.53 (4.12 - 7.43)	215/1510 (14%)	151/1576 (10%)	1.47 (1.19 - 1.81)
Received advice for child nutrition	1432/2562 (56%)	1175/2560 (46%)	1.22 (1.14 - 1.30)	1649/2541 (65%)	1610/2563 (63%)	1.03 (0.99 - 1.08)
Health knowledge						
Mosquito bites are main cause of malaria	2028/2414 (84%)	1990/2421 (82%)	1.02 (0.99 - 1.05)	1942/2078 (93%)	1985/2122 (94%)	1.00 (0.98 - 1.01)
Stop malaria treatment only when dose is completed	2173/2562 (85%)	2116/2560 (83%)	1.03 (1.00 - 1.05)	2332/2541 (92%)	2324/2563 (91%)	1.01 (0.99 - 1.03)
Zinc supplements can treat diarrhea	1275/2562 (50%)	1016/2560 (40%)	1.25 (1.16 - 1.34)	1807/2541 (71%)	1700/2565 (66%)	1.08 (1.03 - 1.12)
Should breastfeed in the hour after birth	2221/2562 (87%)	2199/2560 (86%)	1.01 (0.99 - 1.03)	2027/2541 (80%)	2064/2563 (81%)	0.99 (0.96 - 1.02)
Wait 6m to give fluids other than breastmilk	2124/2562 (83%)	2080/2560 (81%)	1.02 (0.99 - 1.05)	1860/2541 (73%)	1898/2563 (74%)	0.99 (0.95 - 1.03)
Yellowish breastmilk of first few days is good for baby	1603/2562 (63%)	1546/2560 (60%)	1.04 (0.99 - 1.09)	1714/2541 (67%)	1714/2563 (67%)	1.01 (0.97 - 1.06)
A newborn given only breastmilk falls less sick	1544/2562 (60%)	1464/2560 (57%)	1.05 (1.00 - 1.11)	958/2541 (38%)	989/2565 (39%)	0.98 (0.90 - 1.06)
A 1-year-old who eats meat regularly falls less sick	2091/2562 (82%)	2035/2560 (79%)	1.03 (1.00 - 1.06)	1954/2541 (77%)	1970/2565 (77%)	1.00 (0.96 - 1.04)
Prevention						
Treat water before drinking it	2172/2561 (85%)	2087/2555 (82%)	1.04 (1.00 - 1.08)	1817/2536 (72%)	1823/2556 (71%)	1.01 (0.96 - 1.05)
Child slept under a treated bednet last night	2239/3060 (73%)	2196/3092 (71%)	1.03 (0.97 - 1.10)	2596/3039 (85%)	2631/3082 (85%)	1.00 (0.98 - 1.03)
Mother changed diet during pregnancy	1157/1635 (71%)	1073/1661 (65%)	1.10 (1.04 - 1.15)	1287/1797 (72%)	1336/1878 (71%)	1.01 (0.97 - 1.05)

Notes: This Table replicates Table 4 in the main text, excluding replacement households. Data are n (%) from midline sample household survey. Group assignment is based on the baseline cluster of residence of the household. Shares are computed relative to the total valid answers (missing answers are excluded). Follow-up visits after child/infant sick with malaria/diarrhea for children reported sick in the last 3 months. Adjusted risk ratios are computed using a Poisson model, adjusting for stratified randomization. Confidence intervals are constructed using robust standard errors clustered at the cluster (village) level. As a robustness check, Replacement households were excluded from the sample.

Appendix A

Figure A1. Study Location



Note: The map illustrates the 13 districts that are included in the study, as well as the location of the 500 study villages.

Figure A2. Study timeline

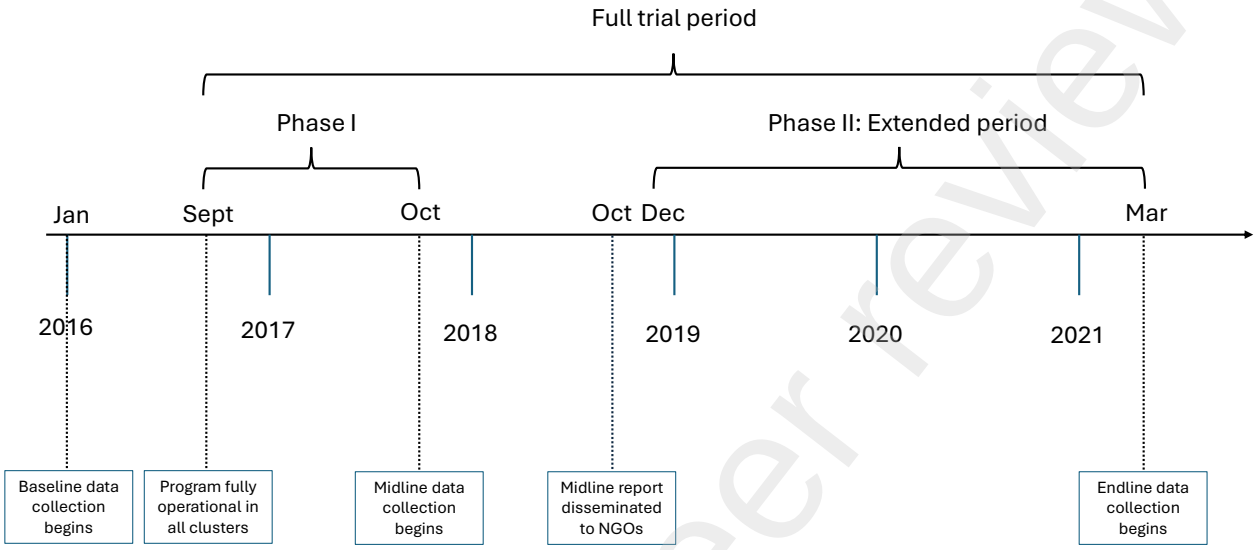
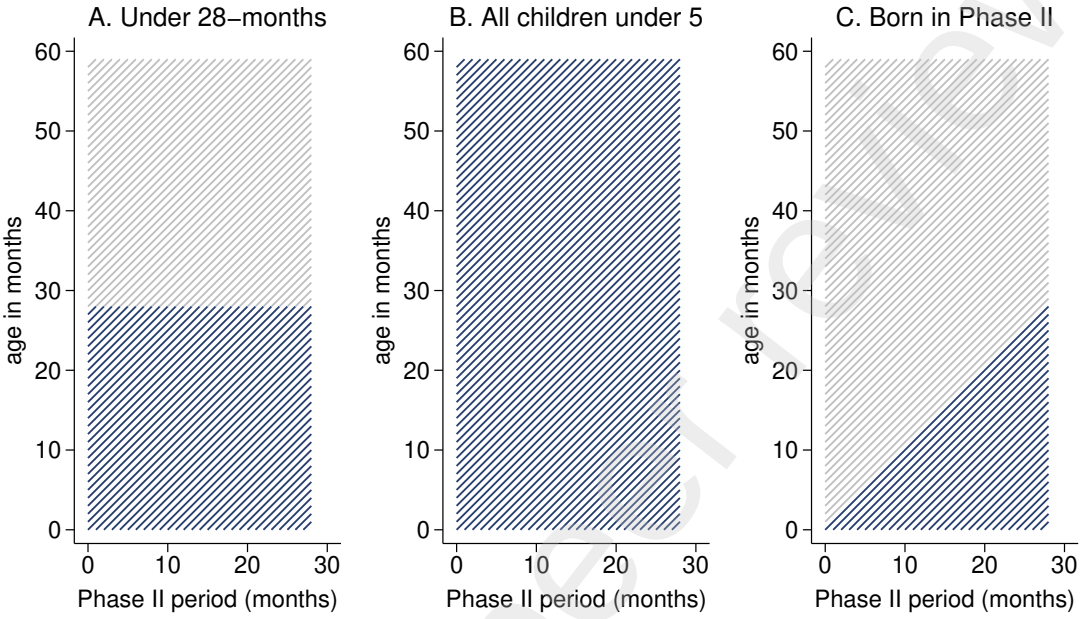


Figure A3. Child mortality samples: Phase II



Note: The x-axis represents the number of months elapsed since the beginning of Phase II, with t=0 marking the starting month. The y-axis corresponds to the age of the child in months. The diagonal lines indicate birth cohorts, while the blue-colored lines represent the child-months included in the sample. In Panel A, exposure months are t=0 to t=27. The sample includes all children born at t∈[-27,-1] and alive at t=0, as well as children born at t∈[0,27]. In Panel B, exposure months are t=0 to t=27. The sample includes all children born at t∈[-58,-1] and alive at t=0, as well as children born at t∈[0,27]. In Panel C, exposure months are t=0 to t=27. The sample includes all children born at t∈[0,27].

Figure A4. NGOs expansion over time (2011-2021)

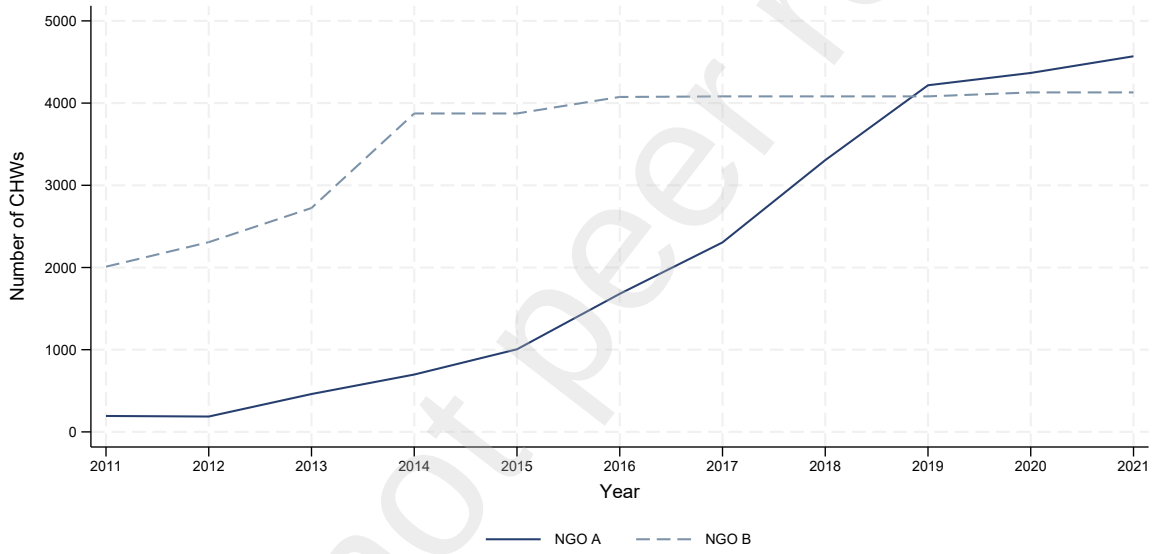


Figure A5. Total Number of active CHWs in study villages

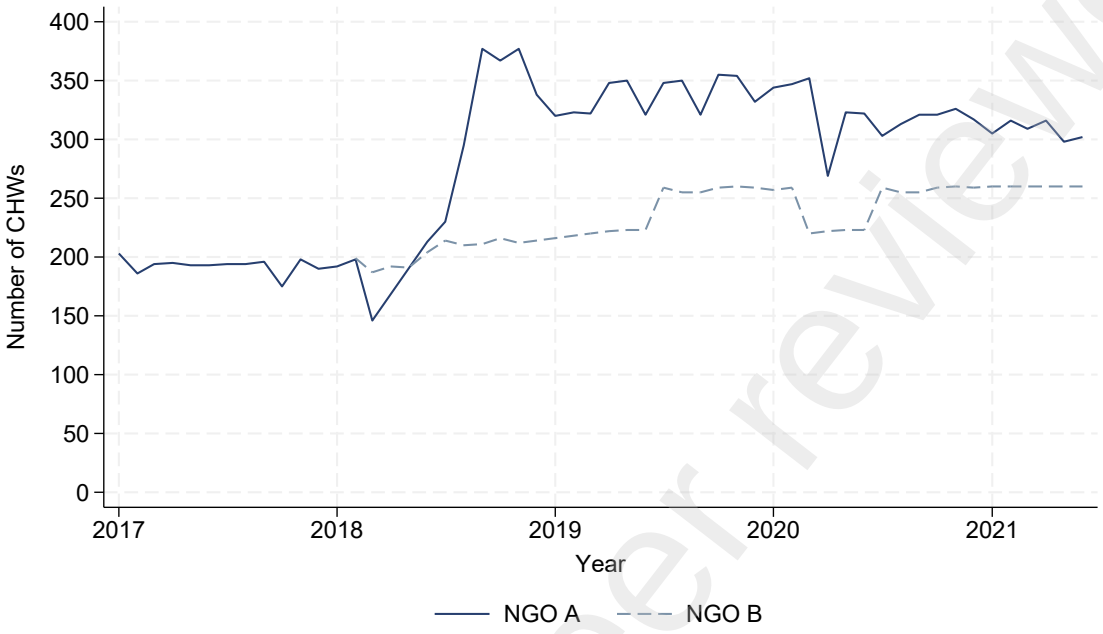
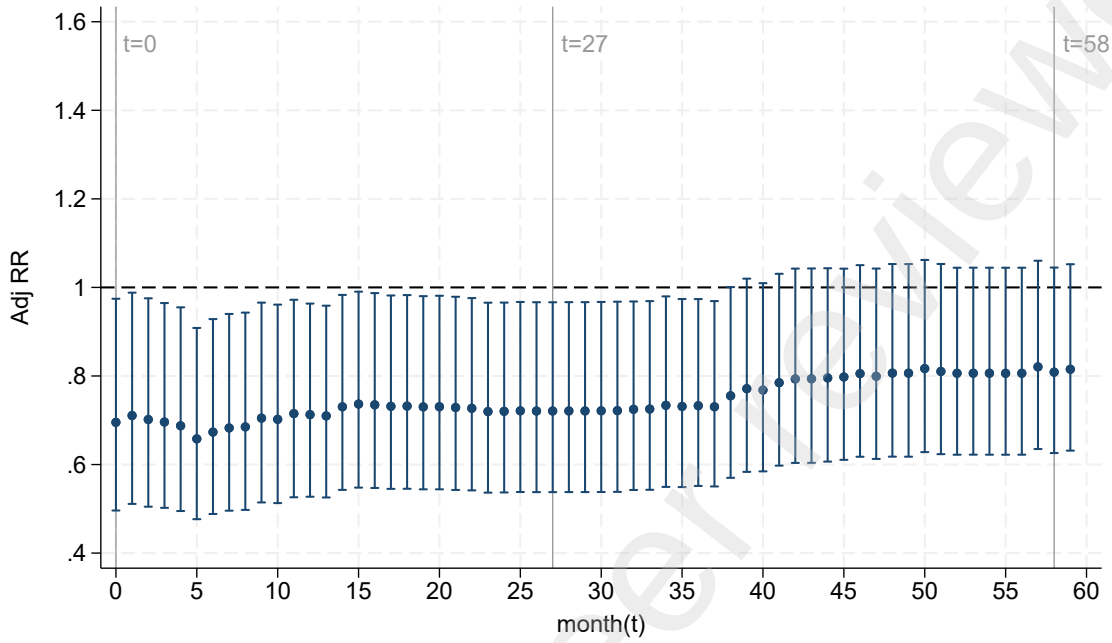
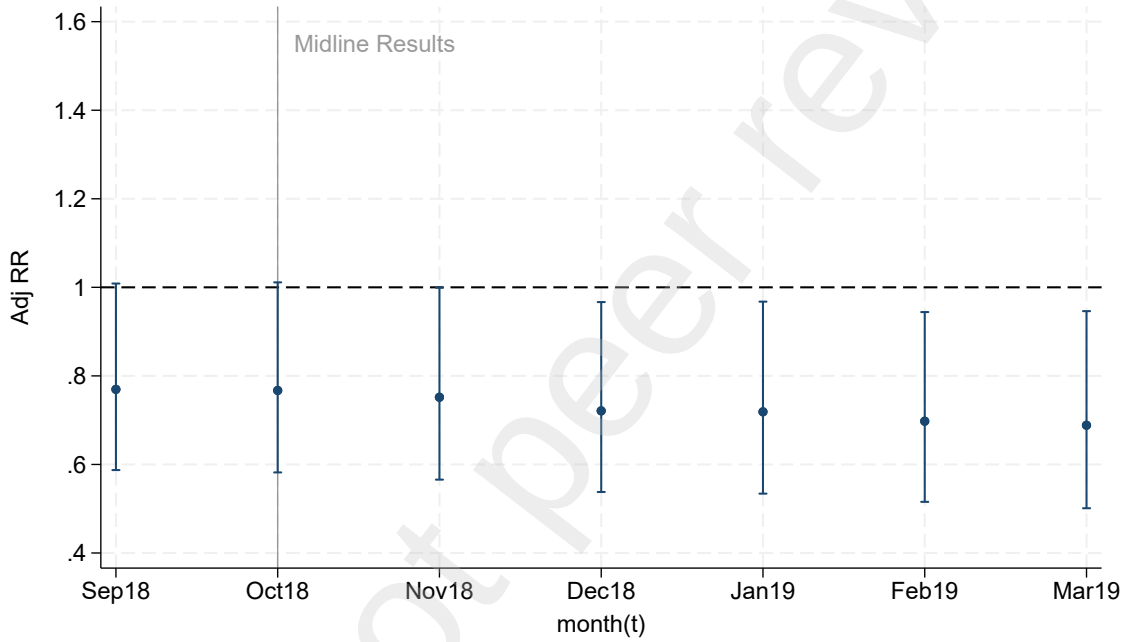


Figure A6. Program A impact on child mortality during phase II for different sample definitions



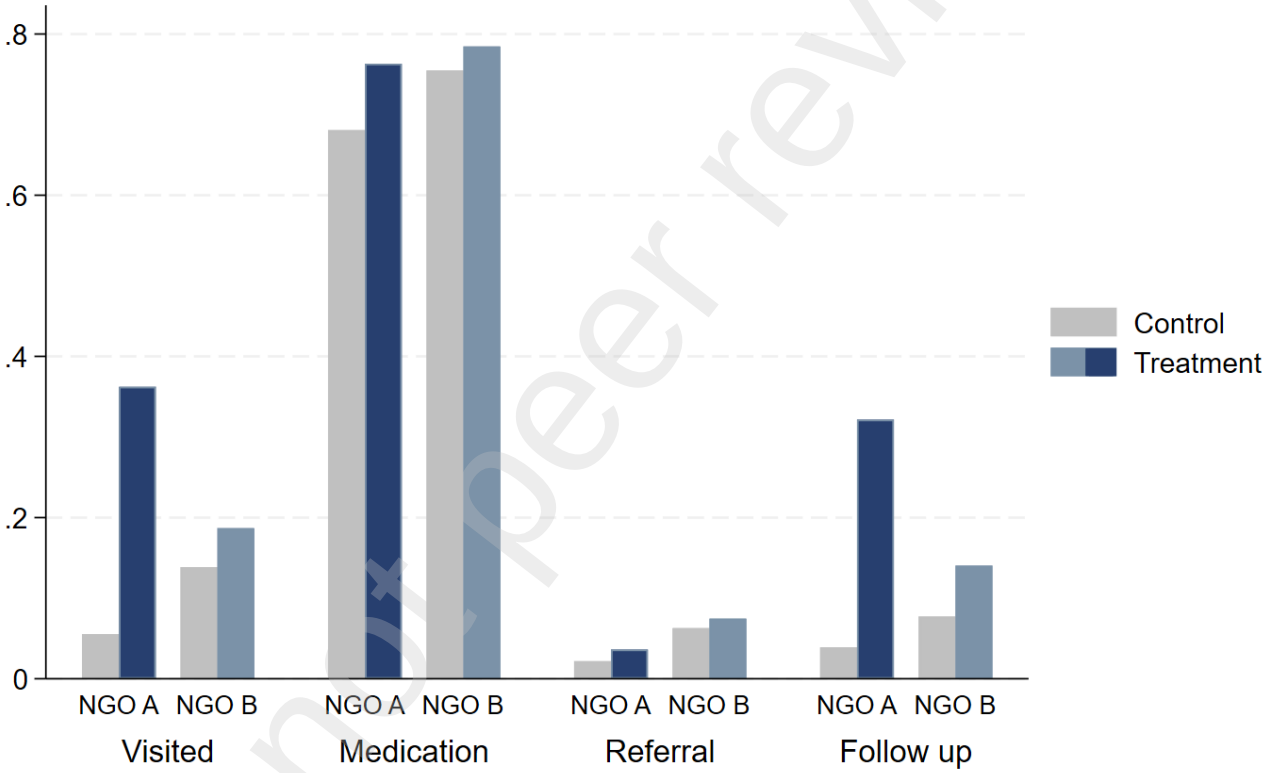
Note: The figure presents the adjusted risk ratios and corresponding 95% confidence intervals (CIs) for child mortality in the NGO A locations, based on different selections of child cohorts included in the mortality analysis. The x-axis represents the birth month of the oldest cohort of children included in the sample, expressed in months relative to the start of Phase II. vertical lines at t=0, t=27, and t=58 define the samples used and reported in Table 3. At t=0, the sample includes children born during Phase II, with the estimate corresponding to Panel C in Table 3. At t=27, the sample includes all children born up to 27 months before Phase II began, with the estimate corresponding to Panel A in Table 3. At t=58, the sample includes all children born up to 58 months before Phase II began, with the estimate corresponding to Panel B in Table 3.

Figure A7. Program A impact on child mortality during phase II for different lengths of Phase II



Note: The figure presents the adjusted risk ratios and corresponding 95% confidence intervals (CIs) for child mortality in the NGO A locations based on the different starting months of the Phase II window. The x-axis represents the starting month of Phase II. The vertical line in October 2018 identifies the time when midline review findings were first discussed with the NGOs. The estimates correspond to the definition of child mortality reported in Panel A in Table 3, i.e. under-X mortality for those exposed to the risk of death during the exposure period, where X depends on the length of Phase II.

Figure A8. Endline Outcomes: Treatment of sick children



Note: The figure displays the proportion of disease cases receiving each type of health service, distinguishing across NGO areas and treatment status. Data is pooled across malaria, diarrhea, and pneumonia cases.

Table A1. Key program adaptations by the two NGOs following midline performance review

Program components	***NGO A***	***NGO B***
Number of CHWs	Number of CHWs increased +60%	Number of CHWs increased +25%
Pricing	Decreased sale price of essential medicines.	No change in prices for essential medicines.
CHW compensation	Increased compensation to \$16–\$20 per CHW per month. Increased share of compensation from performance incentives vis-à-vis compensation from sales.	Small compensation adjustment to \$8 per CHW per month.
Supervision	Strengthened data-driven supportive supervision and coaching through additional training.	No change in supervision

Table A2. Attrition

	Surveyed Households			T vs C
	All	**Treatment**	**Control**	**[p-value]**
Panel A: Baseline				
NGO A	6,286	3,136	3,150	
NGO B	6,329	3,166	3,163	
Total	12,615	6,302	6,313	
Panel B: Endline - Original Sample (% of baseline)				
NGO A	5,082 (80.9%)	2,536 (80.9%)	2,546 (80.8%)	.004 [0.638]
NGO B	5,090 (80.4%)	2,533 (80.0%)	2,557 (80.8%)	-.001 [0.964]
Total	10,172 (80.6%)	5,069 (80.4%)	5,103 (80.8%)	.009 [0.465]

Notes: The last column in Panel B shows the difference in probability of the household being lost at follow-up across treatment and control clusters and the p-value of the test for equality (H0:=probability is equal across treatment and control). The last column in Panel C shows the difference in probability of being a replacement household across treatment and control clusters and the p-value of the test for equality (H0:=probability is equal across treatment and control). The test for equality takes into account district-fixed effects and clustering at the cluster (village) level.

Table A3. Program impact on child mortality - Full Period

	NGO A			NGO B		
	Intervention	**Control**	**Adj RR (95% CI)**	**Intervention**	**Control**	**Adj RR (95% CI)**
Panel A: Under-5 years mortality for those exposed to the risk of death during the exposure period						
Years of exposure to the risk of death	19627	20002		18610	18806	
# child deaths	264	264		161	180	
Mortality rate per 1,000 years of exposure	13.29	13.15	1.02 (0.84 - 1.22)	8.86	9.5	0.90 (0.73 - 1.11)
Panel B: All children born and exposed to the risk of death during the exposure period						
Years of exposure to the risk of death	8120	8366		7897	8144	
# child deaths	190	204		126	143	
Mortality rate per 1,000 years of exposure	23.7	25.27	0.96 (0.77 - 1.19)	16.1	17.47	0.90 (0.71 - 1.14)

Notes: This Table replicates Table A.2, excluding replacement households. Group assignment is based on the baseline cluster of residence of the household. The table report mortality during the full study period, which runs from September 2016 until March 2021. Adjusted risk ratios are computed using a Poisson model, adjusting for stratified randomization. Confidence intervals are constructed using robust standard errors clustered at the cluster (village) level. As a robustness check, Replacement households were excluded from the sample.

Appendix B

Sample Size and Power Calculation

Power calculations were performed using data from the control group in the pilot study conducted in 2011-2013 by the research team in the same setting (Table B1). The original sample size was designed to detect a reduction in under-5 mortality, expressed as number of under-5 deaths per 1000 child-years of exposure to the risk of death under the age of 5. The estimated between-cluster coefficient of variation, k ranged between 0.30 (for neonatal mortality) and 0.43 (for under-5 mortality). As the revised design was expected to make the clusters more similar, k was set equal to 0.30.

Table B2 reports the details of the power computation for both the original and revised design. According to the original design (3 years), a sample size of 250 clusters per arm (500 in total) and 25 households per cluster (12,500 in total) had 80% power to detect a 21% reduction in under-5 mortality at the 0.05 significant level.

Based on the same parameters, the corresponding ex-ante power calculation for the analysis restricted to Phase II indicates that the trial was estimated to detect a 23% reduction in under-5 mortality with the same power and significance level. For each sub-trial (i.e., for the two NGOs separately; each with 250 clusters), restricting the analysis to Phase II, the design was ex-ante powered to detect a 32% reduction in under-5 mortality with 80% power at the 0.05 significance level.

Table B1. Inputs for power computation

Neonatal Deaths:	106
Infant (Under-1) Deaths:	160
Under-5 Deaths:	206
Live Births:	2978
Years of Exposure to risk of death Under-1:	3015
Years of Exposure to risk of death Under-5:	10731
# of months-HHs:	105421

Notes: Data from the control areas in Björkman Nyqvist et al (2019). The reference period is 36 months (2011–2013).

Table B2. Computation details

	Original Design	**Phase II (combined)**	**Phase II (by NGO)**
Mortality Control	0.019	0.019	0.019
Reduction in Mortality	**21%**	**23%**	**32%**
Mortality Intervention	0.015	0.015	0.013
Study-Months	36	28	28
Total Years of Exposure / cluster	77.0	59.9	59.9
Clusters/arm	250	250	125
Total clusters	500	500	250

Notes: Details of the power calculation under the original and revised design. All computations target 80% power and 0.05 significance level and rely on the following assumptions: $k = 0.3$, number of households surveyed at Baseline = 25, attrition by endline = 16%.