Katherine Donato
Adrian Garcia Mosqueira

Power to the people?
A replication study of a community-based monitoring programme in Uganda

August 2016

Replication Paper 11

Public sector management

International Initiative for Impact Evaluation

About 3ie

3ie is an international grant-making NGO promoting evidence-informed development policies and programmes. We are the global leader in funding and producing high-quality evidence of what works, how, why and at what cost. We believe that better and policy-relevant evidence will make development more effective and improve people's lives.

3ie Replication Paper Series

The 3ie Replication Paper Series is designed to be a publication and dissemination outlet for internal replication studies of development impact evaluations. Internal replication studies are those that reanalyse the data from an original paper in order to validate the results. The series seeks to publish replication studies with findings that reinforce an original paper, as well as those that challenge the results of an original paper. To be eligible for submission, a replication study needs to be of a paper in 3ie's online Impact Evaluation Repository and needs to include a pure replication. 3ie invites formal replies from the original authors. These are published on the 3ie website together with the replication study.

The 3ie Replication Programme also includes grant-making windows to fund replication studies of papers identified on the candidate studies list. Requests for proposals are issued one to two times a year. The candidate studies list includes published studies that are considered influential, innovative or counterintuitive. The list is periodically updated based on 3ie staff input and outside suggestions. The aim of the 3ie Replication Programme is to improve the quality of evidence from development impact evaluations for use in policymaking and programme design.

About this report

This paper was funded through 3ie's replication window with generous funding from an anonymous donor. All content is the sole responsibility of the authors and does not represent the opinions of 3ie, its donors or the 3ie Board of Commissioners. Any errors and omissions are the sole responsibility of the authors. Comments and queries should be directed to the corresponding author, Katherine Donato, at katherinedonato@fas.harvard.edu.

Suggested citation: Donato, K and Garcia Mosqueira, A, 2016. *Power to the People?: a replication study of a community-based monitoring programme in Uganda*. 3ie Replication Paper 11. Washington, DC: International Initiative for Impact Evaluation (3ie).

3ie Replication Paper Series executive editor: Annette N Brown
Managing editor: Benjamin DK Wood
Production manager: Brigid Monaghan
Copy editor: Jaime L Jarvis
Proof reader: Yvette Charboneau
Cover design: John F McGill and Akarsh Gupta
Printer: Via Interactive

Power to the people?: a replication study of a community-based monitoring programme in Uganda

Katherine Donato
Harvard University

Adrian Garcia Mosqueira
Harvard University

3ie Replication Paper 11
August 2016

Acknowledgements

We are grateful to 3ie for funding and for technical review and support throughout this study under their Replication Programme. We would also like to thank Martina Björkman and Jakob Svensson for providing the data and code for the original paper and for their feedback on an earlier draft of the replication report. Finally, we thank the internal and external reviewers who provided extremely valuable input on an earlier draft of this report.

Summary

In last few decades, reducing child mortality and improving health outcomes have been pressing objectives in Sub-Saharan Africa and much of the developing world. Although there are a wide range of potentially valuable interventions designed to combat these problems, there is mounting evidence that issues such as absenteeism and leakage of public funds can hinder the effectiveness of these interventions. Björkman and Svensson (2009) show that a relatively simple intervention – providing community-level health service delivery information and guidance on community-based monitoring – dramatically improved health outcomes, even rivalling some of the most effective health interventions to date.

In this paper we conduct a pure replication of Björkman and Svensson's study and extend the original analysis. In the pure replication, we follow the authors' models and specifications and successfully verify the authors' results. In our additional analyses, we extend the community pretreatment balance checks to include household-level checks and ease concerns that pretreatment household imbalances, especially on wealth measures, were driving the results. We also re-examine the statistically significant effect of the treatment on child vaccination rates, finding that the treatment group had a higher vaccination rate at baseline, before the intervention. This calls into question the impact of the community-based monitoring treatment on this outcome. After controlling for the effect of a participating community-based organisation's pre-existing presence in the study areas, the measured programme impacts on under-5 mortality and weight-for-age z-scores lose some magnitude and significance.

Overall, our analysis of the evidence clearly supports the finding that the intervention modified healthcare provider behaviours and utilisation. These improvements suggest that community-based monitoring programmes have the potential to improve their first stage goals of improving provider behaviour and healthcare utilisation, and are worthwhile for policymakers to continue pursuing. Our analysis of the programme's impact on health outcomes does not rule out meaningful results but does suggest that these results may be less robust.

Contents

List of figures and tables

Appendix figures and tables

Abbreviations and acronyms

3ie	International Initiative for Impact Evaluation
BCG	Bacille Calmette-Guérin
DD	Difference in differences
CBM	Community-based monitoring
CBO	Community-based organisation
DPT	Diphtheria, pertussis and tetanus
HUMC	Health unit management committee
OLS	Ordinary least squares

1. Introduction

Seven million children under 5 died in 2011 alone, but this number could have been halved with better access to inexpensive and simple interventions (World Health Organization 2013). While the total number of under-5 deaths has steadily declined over the past two decades, the proportion preventable by these kinds of simple interventions remains stubbornly high (World Bank 2013). The development literature often focuses on the introduction of medical treatments or external products to combat common causes of child mortality including malaria, malnutrition and pneumonia, with mixed results.[1] Although these types of interventions can certainly be effective, there is evidence that issues such as absenteeism of key public service employees – including teachers and health workers – and leakage of public funds significantly influence health outcomes (e.g. Chaudhury *et al.* 2006; Reinikka & Svensson 2004). This suggests that proper administration and oversight of health workers and existing resources may lead to more effective delivery of medical services, yielding subsequent improvements in health outcomes.

Through a randomised controlled trial in nine districts of Uganda, Björkman and Svensson (2009) show that a relatively simple intervention – providing community-level health service delivery information and guidance on community-based monitoring (CBM) – dramatically improved health outcomes, even rivalling some of the most effective health interventions to date (e.g. Jones *et al.* 2003; Kidane & Morrow 2000). Over the course of three meetings, local health service providers informed treatment communities about their rights and entitlements and gave them report cards showing quantitative performance measures for each provider in their community, such as drug availability and interactions with patients. The report cards included performance comparisons between local providers and other facilities, as well as the national standards for these healthcare measures. Using these community-specific parameters, community members and service providers jointly developed action plans for accountably monitoring and addressing their healthcare service deficiencies. Six months later, the implementing organisations returned to conduct a midterm review and a one-day meeting similar to the first three meetings. Where they found improvements, they developed strategies for sustainability.

Björkman and Svensson (2009) find that treatment communities saw significant improvements in primary care provision, utilisation and some health outcomes a year after the intervention. Specifically, they find that treatment facilities were significantly more likely have better organised and informative clinics, installing suggestion boxes, posters on patients' rights and numbered waiting cards. Providers in these facilities served more patients, absenteeism rates fell and average patient waiting times fell, even with an increased caseload. These improvements in service quality and utilisation led to a 33 per cent reduction in under-5 mortality – a significant and very large effect – and moderate improvements in child weight-for-age z-scores.

[1] For example, Miguel and Kremer (2004) find that school-wide introduction of deworming pills increased school attendance by one quarter. Cohen and Dupas (2010) find that women who were randomly assigned to receive free nets were more likely to use the nets than those who paid a subsidised price, making the free nets a more cost-effective lifesaving intervention.

2. Background

Björkman and Svensson (2009) contribute to the growing literature on the importance of monitoring of public services and on community involvement in public service delivery. Over the last few decades, this push towards greater community involvement in development projects has become central to much development policy, drawing more than $85 billion in funding since 2000 from the World Bank alone (Mansuri & Rao 2013).

CBM-style programmes are broadly motivated by the idea that in circumstances where the state's ability to hold public service workers accountable is weak, CBM interventions provide a relatively low-cost means of improving service quality and efficiency (Barr *et al.* 2012; Banerjee *et al.* 2004). Pervasive issues such as corruption, absenteeism, poor adherence to clinical procedures, patient fee irregularities and prescription drug leakages (McPake *et al.* 1999) suggest that the returns from monitoring health service providers could be especially large in Uganda and similar contexts, given the absence of robust oversight institutions.

The first major push towards 'participatory development' occurred in the 1950s and 1960s, driven by the United States Agency for International Development and other donors (Mansuri & Rao 2013). Interest in these programmes declined in the 1970s after a series of failures and the realisation that understanding of how to implement such projects was largely lacking. Around that time, major funders reverted their focus to large-scale investments in the agricultural and industrial sectors. By the mid-1980s, interest in participatory development programmes that leveraged communities' 'social capital' rebounded and criticisms of 'top-down' programmes disempowering the poor resurfaced (ibid.).

Despite this renewed interest in more 'bottom-up' programmes, there has been little effort to systematically analyse what makes some participatory initiatives successful, while others appear to have no effect (ibid.). The mechanisms driving individual policies' effectiveness remain poorly understood (Barr *et al.* 2012). There are prominent examples of CBM-style programmes that appear to have had little measurable effect (e.g. Olken 2007; De Laat *et al.* 2008).

Part of the challenge in designing effective CBM-style interventions is the array of potential options and resulting trade-offs. For example, engaging small groups can be effective because they can be coordinated more easily, but large groups may make more sense if the desired outcome would be enjoyed by a broader group. Moreover, complaints from a small set of individuals may be less effective at inducing change than protests from a large group (Banerjee *et al.* 2004).

There is also a range of options for what such an intervention could actually incorporate. Some CBM-style interventions are relatively simple, offering only information about individuals' rights and leaving them to establish mechanisms for attaining those rights. Others are more involved, including training on how to organise demonstrations or other public action (Banerjee *et al.* 2004). Findings thus far do not include a clear message about this question. In the context of public schools, Barr *et al.* (2012) find that a CBM intervention (providing a school monitoring committee with a scorecard developed by outside groups) was much less effective than a more participatory version, and Banerjee *et al.* (2010) find no impact of a range of CBM-style interventions on teacher effort or student performance.

The Björkman and Svensson paper is influential on another key dimension of CBM-programme design: the importance of information dissemination. The authors argue that the health improvements resulting from this CBM intervention stemmed from the elimination of information gaps among recipients of healthcare services. When these gaps closed, community members were empowered to demand better services that improved health outcomes. Supporting this point, Banerjee, Deaton and Duflo (2004) attribute lack of information as the key reason why their intervention on health worker monitoring in India had no positive results. In a later paper, Björkman, de Walque and Svensson (2014) reinforce the importance of the informational feature of CBM programmes.

The Björkman and Svensson paper has been widely influential in academic circles, as can be seen by its numerous citations since it was published in 2009. The relative simplicity of the intervention also makes the results appealing to policymakers with limited resources, underscoring the importance of exploring the robustness of these results before adopting them in broader contexts. Given that administrative and oversight issues are not unique to the Ugandan context – many other countries in the developing world face similar constraints (e.g. Lindelow & Serneels 2006) – replication of the original authors' results is an important step in promoting or deciding to alter similar programmes.

Verification of Björkman and Svensson's results would lend credence to the importance of CBM in the literature and provide more motivation to expand this important approach in other contexts, including more broadly in Uganda. On the other hand, challenges in verifying the results would suggest a more cautious expansion of the approach and prompt attempts to learn what dimensions could be altered to make it more effective.

Our main replication questions, detailed in a published analysis plan (Donato & Garcia Mosqueira 2013), were designed to help policymakers determine whether to pursue CBM in their countries and what dimensions of the programme to emphasise. Unless otherwise specified, the analyses presented below follow the analysis plan; when analyses differ from those that are pre-specified, we clearly note this and provide justification.

3. Pure replication

Our replication begins by exactly reconstructing key results from the Björkman and Svensson paper, while maintaining the authors' definitions and assumptions. The replication focuses on verifying pretreatment balance, measurements of facility procedures changes, utilisation and objective health outcomes. We also seek to validate the original authors' robustness checks, particularly those that rule out alternative explanations of causality.

The authors' analysis follows three main statistical frameworks, depending on the availability of the data collected in 2004 and 2006 from the 50 study health facilities and about 5,000 households, identified through a stratified random sample of households in the facilities' catchment areas.

To understand the impact of the intervention, we and the authors estimate:

$$y_{ijd} = \alpha + \mathrm{B}T_{jd} + X_{jd}\Pi + \theta_d + \epsilon_{ijd}, \qquad (1)$$

where y_{ijd} is the outcome of household i (if appropriate) in community (health facility) j in district d. T_{jd} is an indicator for being a treatment community, X_{jd} is a vector of pre-intervention, facility-specific covariates, including number of villages in the catchment area, number of days without electricity at the facility in the past month, an indicator for whether the facility has a separate maternity unit, the distance to the nearest public health provider, the number of staff at the facility with less than advanced A-level education, an indicator for whether the staff at the facility could drink water safely from the water source, and the facility's average monthly supply of quinine. Lastly, θ_d are district fixed effects and ϵ_{ijd} is the error term.

When data are available from both the baseline (2004) and endline (2006), we conduct a difference-in-difference analysis, estimating:

$$y_{ijt} = \gamma POST_t + \beta_{DD}(T_j * POST_t) + \mu_j + \epsilon_{ijt}, \qquad (2)$$

where $POST_t$ is an indicator for endline, μ_j is a community-specific fixed effect and β_{DD} is the estimated programme impact.

When data are available on a family of K outcomes, we estimate a seemingly unrelated regression system:

$$Y = [I_K \otimes (TX)]\theta + \nu, \qquad (3)$$

where I_K is a K by K identity matrix. We then estimate average standardised treatment effects.

3.1 Note on standardised effects

When sharing the data and Stata code with us, the authors pointed out that they had identified an error in their calculation of the standardised treatment effects that are reported throughout the paper. The authors informed us that they had inadvertently used the treatment group standard deviations, rather than the control group standard deviations, leading to incorrectly high standardised treatment effects.[2] They had shared this error with the editors of the publishing journal, and all agreed that they would not change the published version, since the statistical significance and direction of the effects remains and the original paper did not discuss quantitative significance of the estimates.

As a result of the original inadvertent error, the magnitudes of our reported standardised effects differ significantly from the original authors', but the directions always match, and the estimates are nearly always of the same significance level. We agree with the authors that the importance of the results depends largely on the direction and statistical significance of the standardised effects, rather than the magnitude, and that the broader findings stands.

[2] Our attempts to replicate the original standardised treatment effects results were unsuccessful. We did not find that switching to the treatment group standard deviations resulted in the originally reported standardised treatment effects.

3.2 Data and .do files

The original authors shared data collected in the household and facility baseline and endline surveys. We did not find any errors in the data (cases where the values were out of range), suggesting that some basic data cleaning had occurred before we received the datasets. In many cases, the variables used in the analyses below were simple indicators based on individual survey questions. Occasionally, variables were constructed using multiple survey question responses. In the online replication documents, which are available on 3ie's Replication Dataverse, we include (1) a description of all of the variables that were used in the tables below; (2) descriptions of how each variable was constructed, which indicates when an outcome was collected during both survey rounds but data were only used from one round in the main paper; and (3) our .do files to produce the tables below (Donato & Garcia Mosqueira 2016).

3.3 Pretreatment balance

Our pure replication begins with Table 1 in order to verify that randomisation was successful on relevant observable facility and catchment area characteristics. This gives us more confidence that the analysis has the potential to truly represent intervention effects and not pretreatment differences between the two comparison groups.

The original results of Table 1 and our replication are presented in our Table 1. Our estimates for the first panel, describing key characteristics of the study health facilities at baseline in 2004, match those presented by the original authors. Facilities appear similar on characteristics such as number of patients seen, number of households in the catchment area and characteristics of the facility capacities. Although our estimates in the second panel, average standardised pre-treatment effects for several families of outcomes, are different from the originally reported values, this was to be expected, as we discuss in section 3.1. As in the original paper, we cannot reject the null that there are no imbalances in these families of outcomes.[3]

Column 5 reports the average standardised effect of the four monitoring tools from the health facilities. Because no control group facility had a suggestion box, it is only possible to calculate the average standardised effect by either excluding the 'suggestion box' outcome or combining it with the second outcome, 'numbered waiting cards' (as the authors suggested in their notes to us). We calculate the average standardised effect using both methods (reporting the former here). Although the effect is smaller than originally published, for the reasons discussed above (see section 3.1), it remains positive and highly significant.

[3] For most families of outcomes in table 1, panel 2, the unit of analysis matches the data source (e.g. household responses are analysed at the household level). The authors' analysis of user charges, measured at the household level, breaks from this pattern, and is instead analysed after collapsing by facility. If the analysis is done at the individual level, the average standardised effect is significant at the 10 per cent level, with an estimate of –0.26 (0.13).

Table 1: Pretreatment facility and catchment area characteristics and average standardised effects

Variables	Treatment group		Control group		Difference	
	(1)	(2)	(3)	(4)	(5)	(6)
Key characteristics						
Outpatient care	593	593	675	675	-82	-82
	(75)	(75)	(57)	(57)	(94)	(94)
Delivery	10.3	10.3	7.5	7.5	2.8	2.8
	(2.2)	(2.2)	(1.4)	(1.4)	(2.6)	(2.6)
No. of households in catchment area	2,140	2,140	2,224	2,224	-84.4	-84.4
	(185)	(185)	(204)	(204)	(276)	(276)
No. of households per village	93.9	93.9	95.3	95.3	-1.42	-1.42
	(5.27)	(5.27)	(6.32)	(6.32)	(8.23)	(8.23)
Drank safely today	0.40	0.40	0.32	0.32	0.08	0.08
	(0.10)	(0.10)	(0.10)	(0.10)	(0.14)	(0.14)
No. of days without electricity in past month	18.3	18.3	20.4	20.4	-2.12	-2.12
	(2.95)	(2.95)	(2.90)	(2.90)	(4.14)	(4.14)
Average standardised pretreatment effects						
Utilisation					0.11	0.066
					(0.77)	(0.27)
Utilisation pattern					-0.48	-0.040
					(0.33)	(0.028)
Quality measures					-0.35	-0.036
					(0.84)	(0.091)
Catchment area statistics					0.11	0.09
					(0.66)	(0.20)
Health facility characteristics					0.14	0.073
					(0.31)	(0.092)
Citizen perceptions					0.37	0.011
					(0.67)	(0.055)
Supply of drugs					0.73	0.17
					(0.83)	(0.26)
User charges					-0.65	-0.46
					(0.63)	(0.32)

Note: Values from original analysis are in grey; replicated values are in white. Key characteristics are catchment area/health facility averages for treatment and control group and difference in averages. Robust standard errors are in parentheses. Description of variables: 'Outpatient care' is average number of patients visiting the facility per month for outpatient care. 'Delivery' is the average number of deliveries at the facility per month. 'Number of households in catchment area' and 'number of households per village' are based on census data and maps from the Uganda Bureau of Statistics. 'Drank safely today' is an indicator variable for whether health facility staff could safely drink from the water source at the time of the preintervention survey. 'Number of days without electricity in the past month' is measured out of 31 days. 'Average standardised pretreatment effects' are derived by estimating equation (3) on each household's outcomes. 'Utilisation' summarises outpatients and deliveries. 'Utilisation pattern' summarises the households' reported utilisation locations. 'Quality measures' summarises households' reports of waiting time and whether equipment was used. 'Catchment area statistics' summarise four measures of the health facility catchment area population. 'Health facility characteristics' summarise 10 measures, including staff education and distance to the nearest public health provider. 'Citizen perceptions' summarises four subjective measures of households' experience during their last visit to the project health facility. 'Supply of drugs' summarises the health facility's average free receipt of five key drugs. 'User charges' summarises households' reports of paying for four key health services.

The last two columns of Table 2 analyse data from the follow-up household surveys based on equation (1). In columns 6 and 7, the findings are consistent with the results presented by the original authors: households in treatment communities were more likely to have discussed the functioning of the health facility during community-level meetings and to have received information about the Health Unit Management Committee (HUMC) and its roles and responsibilities.

Table 2: Programme impact on monitoring and information

Dependent variable	Suggestion box	Numbered waiting cards	Poster informing free services	Poster on patients' rights	Average standardised effect	Discuss facility in local council meetings	Received information about HUMC
Specification	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Programme impact	0.32***	0.16*	0.27***	0.14	2.55***	0.13***	0.04***
	(0.08)	(0.09)	(0.09)	(0.10)	(0.55)	(0.03)	(0.01)
Mean control group	0	0.04	0.12	0.12	-	0.33	0.08
Observations	50	50	50	50	50	3,119	4,996
Programme impact	0.32***	0.16*	0.27***	0.14	0.68***	0.13***	0.04***
	(0.08)	(0.09)	(0.09)	(0.10)	(0.25)	(0.03)	(0.01)
Mean control group	0	0.04	0.12	0.12	-	0.33	0.08
Observations	50	50	50	50	50	3,119	4,996

Note: Values from original analysis are in grey; replicated values are in white. Robust standard errors in parentheses, clustered by catchment areas in columns 6 and 7. Point estimates, standard errors and average standardised effect (columns 1–5) are derived from equation (3). Programme impact measures the coefficient on the assignment to treatment indicator. Outcome measures in columns 1–4 are based on data collected through visual checks by the enumerators during the post-intervention facility survey. Outcome measures in columns 6 and 7 are from the post-intervention household survey. The estimated equations all include district fixed effects and the following baseline covariates: number of villages in catchment area, number of days without electricity in the past month, 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, indicator variable for whether the staff could safely drink from the water source, and average monthly supply of quinine. Specification: (1) indicator variable for whether the health facility has a suggestion box for complaints and recommendations; (2) indicator variable for whether the facility has numbered waiting cards for its patients; (3) indicator variable for whether the facility has a poster informing about free health services; (4) indicator variable for whether the facility has a poster on patients' rights and obligations; (5) average standardised effect of the estimates in columns 1–4; (6) indicator variable for whether the household discussed the functioning of the health facility at a local council meeting during the past year; (7) indicator variable for whether the household has received information about the HUMC's roles and responsibilities. *Significant at 10%. **Significant at 5%. ***Significant at 1%.

3.4 Treatment practices

The results in Table 2 suggest that monitoring and information provision at treatment health facilities improved as a result of the CBM intervention, and Table 3 offers further evidence that the intervention influenced the health workers' actions and performance during the provision of care and management of the health facility. Rows 1 and 3 are based on equation (2), and the remaining rows are based on the ordinary least squares (OLS) model in equation (1). Our analysis shows nearly identical results compared to the original paper. The first row shows that providers in treatment facilities were more likely to use some piece of equipment (e.g. a thermometer) during the examination. The third row shows that waiting times for patients were reduced by about 10 minutes at treatment facilities, compared with an average of more than 2 hours. The fifth row shows that the effect on the absence rate (the ratio of workers not physically present at the time of the follow-up survey to workers identified during the baseline preintervention survey) falls by 13 percentage points. Row 6 shows statistically significant improvements on a composite score of the cleanliness of the facility and rows 7 and 8 (information provided to patients on the dangers of self-treatment and the importance of family planning). The last row shows that although treatment facilities

provided care for more patients, they were significantly less likely to have drug stock-outs, suggesting that there were more drug leakages in the control group facilities.

Table 3: Programme impact on treatment practices and management

Spec.	Dependent variable	Model	Programme impact	2005	Mean control group 2005	Obs.
(1)	Equipment used	DD	0.08**	–0.07***	0.41	5,280
			(0.03)	(0.02)		
(1)	Equipment used	DD	0.08**	–0.07***	0.41	5,280
			(0.03)	(0.02)		
(2)	Equipment used	OLS	0.01		0.41	2,758
			(0.02)			
(2)	Equipment used	OLS	0.01		0.41	2,758
			(0.03)			
(3)	Waiting time	DD	–12.3*	–12.4**	131	6,602
			(7.1)	(5.2)		
(3)	Waiting time	DD	–14.10*	–9.98*	133	5,148
			(7.69)	(5.22)		
(4)	Waiting time	OLS	–5.16		131	3,426
			(5.51)			
(4)	Waiting time	OLS	–4.67		133	2,694
			(6.79)			
(5)	Absence rate	OLS	–0.13**		0.47	46
			(0.06)			
(5)	Absence rate	OLS	–0.13**		0.47	46
			(0.06)			
(6)	Management of clinic	OLS	1.20***		-0.49	50
			(0.33)			
(6)	Management of clinic	OLS	1.20***		-0.49	50
			(0.33)			
(7)	Health information	OLS	0.07***		0.32	4,996
			(0.02)			
(7)	Health information	OLS	0.07***		0.32	4,996
			(0.02)			
(8)	Importance of family planning	OLS	0.06***		0.31	4,996
			(0.02)			
(8)	Importance of family planning	OLS	0.06***		0.31	4,996
			(0.02)			
(9)	Stock-outs	OLS	–0.15**		0.50	42
			(0.07)			
(9)	Stock-outs	OLS	–0.15**		0.50	42
			(0.07)			

Note: Values from original analysis are in grey; replicated values are in white. Each row is based on a separate regression. The difference-in-differences (DD) model is from equation (2). The OLS model is from equation (1), with district fixed effects and baseline covariates as listed in Table 2. Robust standard errors, clustered by catchment areas, are in columns 1–4 and 7–8, in parentheses. Specifications: (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 citizen left the facility and the time the citizen arrived at the facility, minus examination time; (5) ratio of workers not physically present at the time of the post-intervention survey to workers employed preintervention; (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 enumerators ranked each condition from 1 (dirty) to 3 (clean); (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 months in 2005 in which stock cards indicated no availability of drugs. *Significant at 10%. **Significant at 5%. ***Significant at 1%.

In Table 4 we report the results of estimating equation (3) on an indicator for whether the child has received the appropriate immunisations by age group, looking at polio; diphtheria, pertussis and tetanus (DPT); bacille Calmette-Guérin (BCG); and measles vaccines, as well as vitamin A supplements. For reasons discussed in section 3.1, the magnitude of the effect sizes is smaller in our estimates than the original paper, but the direction is consistent and statistical significance is generally unchanged. The improvements in immunisation rates appear relatively strong for younger cohorts.

Table 4: Programme impact on immunisation

Group	Newborn	Under 1 year	1 year old	2 years old	3 years old	4 years old
Specification:	(1)	(2)	(3)	(4)	(5)	(6)
Average standardised effect	1.3*	1.44*	1.24**	0.72	2.01***	0.86
	(0.70)	(0.72)	(0.63)	(0.58)	(0.67)	(0.80)
Observations	173	929	940	951	1,110	526
Average standardised effect	0.35**	0.09*	0.070*	0.023	0.11***	0.058
	(0.16)	(0.052)	(0.039)	(0.032)	(0.035)	(0.067)
Observations	173	929	940	951	1,110	526

Note: Values from original analysis are in grey; replicated values are in white. Average standardised effects are derived from equation (3), with the dependent variables being indicator variables for whether the child has received the required dose(s) of measles, DPT, BCG and polio vaccines and vitamin A supplementation, respectively, and with district fixed effects and baseline covariates listed in Table 2 included. Robust standard errors clustered by catchment areas in parentheses. Groups: (1) children under 3 months; (2) children 0–12 months; (3) children 13–24 months; (4) children 25–36 months; (5) children 37–48 months; (6) children 49–60 months. *Significant at 10%. **Significant at 5%. ***Significant at 1%.

3.5 Utilisation

Tables 2–4 present evidence that communities in treatment areas increased monitoring of their health facilities, that health facilities implemented measures to allow them to more closely gauge the communities' assessments and that providers improved treatment. Table 5 presents evidence on the quantity of services provided. The first panel considers the outcomes cross-sectionally at endline and reports coefficient estimates from equation (1). The second panel leverages panel data when available, reporting estimates based on equation (2). Our estimates of programme impact match nearly identically, with the exception of the average standardised effects, which differ for the reasons noted in section 3.1 but maintain the direction and significance of the original estimates. Outpatient and delivery utilisation, measured using facility register data, increased in treatment communities. Antenatal and family planning visits may have increased but are imprecisely estimated. The last three columns analyse household-level data, finding that households in treatment communities increased their use of the treatment facility and reduced use of traditional healers and self-treatment.

Table 5: Programme impact on utilisation/coverage

Dependent variable	Outpatients	Delivery	Antenatal	Family planning	Avg. std. effect	Use of project facility	Use of self-treatment/ traditional healers	Avg. std. effect
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
A. Cross-sectional data								
Programme impact	130.2** (60.8)	5.3** (2.1)	15.0 (11.2)	3.4 (3.2)	1.75*** (0.63)	0.026* (0.016)	-0.014 (0.011)	1.43* (0.87)
Observations	50	50	50	50	50	50	50	50
Programme impact	130.3** (60.8)	5.3** (2.1)	15.0 (11.2)	3.4 (3.2)	0.46*** (0.16)	0.028* (0.016)	–0.015 (0.012)	0.24* (0.14)
Observations	50	50	50	50	50	50	50	50
	(9)	(10)			(11)	(12)	(13)	(14)
B. Panel data								
Programme impact	189.1*** (67.2)	3.48* (1.96)			2.30*** (0.69)	0.031* (0.017)	–0.046** (0.021)	1.96** (0.89)
Observations	100	100			100	100	100	100
Programme impact	189.2*** (67.2)	3.48* (1.96)			0.76*** (0.16)	0.033* (0.017)	–0.046** (0.022)	0.44*** (0.14)
Observations	100	100			100	100	100	100
Mean control group 2005	661	9.2	78.9	15.2	-	0.24	0.36	-
Mean control group 2005	661	9.2	78.9	15.2	-	0.24	0.36	-

Note: Values from original analysis are in grey; replicated values are in white. Panel A reports programme 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 programme impact estimates from difference-in-differences models with robust standard errors clustered by facility in parentheses. Point estimates, standard errors, and average standardised effects in specifications 1–5, 6–8, 9–11 and 12–13 are derived from equation (3). Specifications: First column is average number of patients visiting the facility per month for outpatient care; second column is average number of deliveries at the facility per month; third column is average number of antenatal visits at the facility per month; fourth column is average number of family planning visits at the facility per month; fifth column is average standardised effect of estimates in specifications 1–4 and 9–10, respectively; sixth column is the share of visits to the project facility of all health visits, averaged over the catchment area; seventh column is the share of visits to traditional healers and self-treatment of all health visits, averaged over the catchment area; eighth column is average standardised effect of estimates in specifications 6–7 and 12–13, respectively, reversing the sign of use of self-treatment/traditional healers. *Significant at 10% level. **Significant at 5% level. ***Significant at 1% level.

3.6 Health outcomes

Given the evidence presented above, which shows that processes, treatment practices and utilisation improved in the treatment communities, there are several channels through which health outcomes could plausibly improve. For example, changing utilisation patterns (from self-treatment or traditional healers to public facilities) or improving care at the time of utilisation could each improve health outcomes in the community. Table 6 presents estimates of the programme's impact on several health outcomes, found through estimating equation (1) on cross-sectional data collected a year after the intervention was introduced.

Table 6: Programme impact on health outcomes

Dependent variable	Births		Pregnancies		Under-5 mortality rate		Child death		Weight-for-age z-score			
Specification:	(1)	(1)	(2)	(2)	(3)	(3)	(4)	(4)	(5)	(5)	(6)	(6)
Programme impact	–0.016	–0.016	–0.03**	–0.03**	–49.9*	–49.9*			0.14*	0.14*	0.14**	0.14**
	(0.013)	(0.014)	(0.014)	(0.014)	(26.9)	(26.9)			(0.07)	(0.08)	(0.07)	(0.07)
Child age (log)											–1.27***	–1.27***
											(0.07)	(0.07)
Female											0.27***	0.27***
											(0.09)	(0.09)
Programme impact x year of birth 2005							–0.026**	–0.026**				
							(0.013)	(0.013)				
Programme impact x year of birth 2004							–0.019**	–0.019**				
							(0.008)	(0.008)				
Programme impact x year of birth 2003							0.003	0.003				
							(0.009)	(0.009)				
Programme impact x year of birth 2002							0.000	0.000				
							(0.006)	(0.006)				
Programme impact x year of birth 2001							0.002	0.002				
							(0.006)	(0.006)				
Mean control group 2005	0.21	0.22	0.29	0.29	144	144.4	0.029	0.029	–0.71	–0.71	–0.71	–0.71
Observations	4,996	4,996	4,996	4,996	50	50	5,094	5,094	1,135	1,135	1,135	1,135

Note: Values from original analysis are in grey; replicated values are in white. Robust standard errors in parentheses. Estimates from equation (1) with district fixed effects and baseline covariates as listed in Table 2 included. Specification 4 also includes a full set of year-of-birth indicators (3), clustered by catchment area (1–2, 4–6). Specifications: (1) Number of births in the household in 2005; (2) indicator variable for whether any women in the household are or were pregnant in 2005; (3) Under-5 mortality rate is the rate in the community expressed per 1,000 live births (see text for details); (4) indicator variable for child death in 2005; (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). *Significant at 10% level. **Significant at 5% level. ***Significant at 1% level.

Births and pregnancies, the first two columns of Table 6, might be influenced by utilisation of family planning services. In line with the authors' findings, our estimates suggest a negative impact on the number of pregnancies.

Following the authors' approach to measurement of the outcome, our estimates of the treatment effect on under-5 mortality and child death in columns 3 and 4 exactly match the original paper, finding a significant 33 per cent reduction. Our estimates of the programme's impact on weight-for-age, shown in columns 5 and 6, are also identical to the original authors', finding improvements of about 0.14 standard deviations.

3.7 Robustness

Although we explore alternative robustness checks in later sections of the replication, Table 7 addresses the authors' exploration of mechanisms and robustness. The goal is specifically to understand whether the difference between the two treatment groups is larger when there are larger differences in monitoring and information outcomes (proxied by a community monitoring index generated using the six monitoring and information variables in Table 2). This is important because there are plausible drivers of the results other than the CBM. For example, it could be that providers directly respond to information given during the initial meetings (e.g. that a provider's performance is far below her peers'), rather than the monitoring conducted by the community.[4]

Following Kling, Liebman and Katz (2007), the authors tested the impact of the monitoring index by two-stage least squares, instrumenting for the monitoring index using the full set of district-by-treatment interactions and controlling for district fixed effects, under the argument that the estimated impact of the monitoring index is consistent if the monitoring index is the mediating factor between treatment and the observed outcomes (outpatient care, under-5 mortality). In columns 5 and 6, we add an additional control for the staff's knowledge about patients' rights and obligations, which is argued to reflect more of effects due to the information reported (facility/staff performance), rather than direct monitoring. In columns 7 and 8 we add an additional control for the presence of an implementing community-based organisation (CBO) prior to the intervention (which occurred in 64 per cent of treatment communities and half of control communities), both individually and interacted with programme impact.

Our findings, presented below, are consistent with the original authors'. We find that the difference between treatment and control communities is indeed larger when the degree of treatment appears larger, with effect sizes that are fairly consistent across multiple specifications. The results do not appear to be driven by staff's knowledge of patient rights, and prior presence by implementing CBOs appears less important than the programme, with coefficient magnitudes and significance remaining approximately the same after including controls for a CBO's prior presence.

[4] We note that four of the six components of the community monitoring index (suggestion box, numbered waiting cards, poster informing free services and poster on patients' rights) could have been influenced directly through the providers' response to new information, rather than pressure from community members (i.e. more supply-driven than demand-driven). If we form an index from the remaining two components (discuss health facility in local council meetings, receive information about HUMC), the coefficients in columns 1 and 2 become 0.88 (p=0.004) and –0.67 (p=0.03), respectively.

Table 7: Mechanisms and robustness

Dependent variable	Outpatients		Under-5 mortality		Outpatients		Under-5 mortality		Outpatients		Under-5 mortality		Out-patients	Under-5 mortality	Out-patients	Under-5 mortality
Specification:	(1)	(1)	(2)	(2)	(3)	(3)	(4)	(4)	(5)	(5)	(6)	(6)	(7)	(8)	(7)	(8)
Community monitoring index	0.77*** (0.22)	0.77*** (0.22)	–0.43* (0.25)	–0.43* (0.25)	0.86* (0.53)	0.77 (0.48)	–0.43 (0.82)	–0.44 (0.82)	0.77** (0.21)	0.77*** (0.21)	–0.54* (0.30)	–0.54* (0.30)				
Staff's knowledge about patients' rights									–0.01 (0.28)	–0.01 (0.28)	0.47 (0.29)	0.47 (0.30)				
Programme impact					–0.12 (0.66)	–0.12 (0.66)	0.01 (0.88)	0.02 (0.89)					190.5** (92.6)	–41.3 (45.8)	190.6** (92.6)	–41.2 (45.8)
CBO presence													–8.3 (69.4)	–21.0 (37.9)	–8.3 (69.4)	–20.9 (37.9)
Programme impact x CBO presence													–127.9 (126.1)	–4.0 (58.4)	–127.9 (126.2)	–4.1 (58.4)
F-test on programme impact													6.16 (0.05)		6.16 (0.05)	
F-test on CBO presence													0.37 (0.83)		0.37 (0.83)	
F-test on programme impact x CBO presence													1.03 (0.60)		1.03 (0.60)	

Note: Values from original analysis are in grey; replicated values are in white. Columns 1–4 report 2SLS estimates from equation (4) with district-by-treatment interactions as the excluded instruments and district fixed effects and outpatients$_{t-1}$ (in specifications 1 and 3) as controls. The variables in columns 1–6 are expressed in standard deviation units relative to the control group overall standard deviation for each variable. Robust standard errors are in parentheses. F-test statistics (with *p*-values in parentheses) on the excluded instruments. Point estimates in columns 5–6 and standard errors in columns 7–8 are jointly estimated from equation (3). Explanatory variables: 'Community monitoring' is the first component from a principal components analysis of the six monitoring and information proxies presented in Table 2. 'Staff's knowledge about patients' rights' is a measure of the in-charge's knowledge about patients' rights and obligations (Donato & Garcia Mosqueira 2016). 'CBO presence' is an indicator variable for whether a participating CBO had been operating in the community before the intervention. 'F-test on programme impact' (CBO presence) [Programme impact × CBO presence] is the test statistic, with *p*-values in parenthesis, on the test that the coefficients on programme impact (CBO presence) [Programme impact × CBO presence] are jointly 0 in columns 5–6 and 7–8, respectively. *Significant at 10% level. **Significant at 5% level. ***Significant at 1% level.

3.8 Pure replication conclusion

In our pure replication, we maintain the authors' variable definitions and analytical specifications and verify their main results. We first find that pretreatment balance on facility characteristics was satisfied. Additionally, we show that the CBM programme improved monitoring and information provision, treatment practices and management, immunisation rates for under-5 children, utilisation and coverage in treatment communities, and health outcomes, including under-5 mortality and weight-for-age. Finally, to understand what mechanisms drove the results, we verify the authors' results that communities with more monitoring saw greater improvements, with some evidence that this result was driven more by changes in community monitoring/information provision than facility staff's knowledge.

4. Additional analyses

Having replicated the results in the original paper following the authors' original specifications and models, we now show the results of additional analyses. We focus our extension on pretreatment balance on household characteristics, understanding the impact of treatment on child immunisations, the impact of including controls for implementing CBOs' presence prior to the intervention and additional process measures that might help explain what specific changes were made as a result of the intervention. Most of these analyses strengthen the results of the main paper, but the findings on child immunisations and incorporating controls for prior CBO presence add some doubt to the relevant results.

Some of these analyses were *not* pre-specified and only became apparent as important checks once we had the data and questionnaires used by the authors (after we published our replication plan). Whenever an analysis below was not pre-specified, we include a star next to the heading and include our justification for the analysis in the text. These additional analyses are designed to provide more support for the validity of the analyses, to verify results by using alternative specifications and to gain more insights into the specific mechanisms behind the observed treatment effects.

4.1 Pretreatment balance on household characteristics

Although the original authors show that there is pretreatment balance between the treatment and control communities on a number of facility- and community-level factors, they do not report any pretreatment household-level characteristics across treatment versus control communities (or use any household-level controls in the main analyses). Considering that the intervention's effects are thought to be driven by accountability and monitoring *from* the households, we feel it is especially important to verify that this pretreatment balance extends to the households.

As our pre-specified analysis plan indicates, we were especially interested in comparing the wealth of the households in each community. It is possible that the treatment group was relatively wealthy compared with the control group, and that the differences in health outcomes (e.g. under-5 mortality) were driven by wealth effects rather than the intervention. Considering that there are no baseline measures of health, it is important to verify baseline balance on household characteristics that may have a significant impact on health.

A limited set of data was collected on household characteristics. Our analysis of these characteristics reveals that there is balance on measured demographic and socioeconomic factors such as education and house quality (a proxy for wealth). A bar graph showing the specific breakdown of education levels for the respondent (the person in the household who most recently sought medical care) is included in the appendix (figure A1).

Table 8: Pretreatment facility and catchment area characteristics and average standardised effects

Variables	Treatment group	Control group	Difference
Household characteristics			
Number of adults in the household	2.59 (0.03)	2.68 (0.03)	–0.09 (0.10)
Any children under 5 in the household	0.66 (0.01)	0.68 (0.01)	–0.03 (0.019)
Number of children under 5 in the household	0.81 (0.021)	0.85 (0.022)	–0.04 (0.06)
Pregnancy in household in last year	0.37 (0.011)	0.38 (0.011)	–0.01 (0.022)
Respondent has secondary education	0.20 (0.008)	0.17 (0.008)	0.031 (0.024)
Low-quality walls	0.56 (0.01)	0.60 (0.01)	–0.041 (0.06)
Low-quality roof	0.30 (0.009)	0.32 (0.01)	–0.026 (0.087)

Note: Household characteristics are averages for treatment and control group and differences in averages. Robust standard errors, clustered at the facility level, are in parentheses in the difference column. All characteristics were measured during the pre-intervention household survey. Description of variables: 'Number of adults in household' counts the number of adults reported to be in the household by the respondent. 'Any children under 5 in the household' is an indicator for whether there are any children under 5 in the household. 'Number of children under 5 in the household includes all households', regardless of whether the household has any children under 5. 'Pregnancy in the household in last year' is an indicator for whether anyone in the household had been pregnant since January 2003. 'Respondent has secondary education' is an indicator for respondents (the person in the household who visited the project health facility most recently) with secondary education, where the comparison is less than secondary schooling. 'Low-quality walls' is an indicator for having walls made of mud and poles or unburnt bricks, compared with burnt bricks with mud or cement, timber, stones, or cement plastered. 'Low-quality roof' is an indicator for having a thatched roof, compared with an iron roof (old or new) or a tiled roof. *Significant at 10%, ** Significant at 5%, *** Significant at 1%.

4.2 Child immunisations*

In the pure replication, Table 4 shows statistically significant improvements in immunisation rates for newborns, children under 1 year, 1-year-old children, and 3-year-old children. The outcomes in this table are only from data collected at endline, but identical questions were asked during the baseline pre-intervention survey. We therefore sought to examine whether the statistically significantly higher immunisation rates in the treatment group were present before the intervention, which would trigger doubt that the effects shown in Table 4 were due only to the treatment. Since we did not have the data or questionnaires before writing our analysis plan, this was not pre-specified.

We constructed the outcome variables using the identical approach that the authors used, which identifies appropriate dosage of immunisations for measles, DPT, BCG and polio, as

well as vitamin A supplementation. We first present a simple balance table in Table 9 (analogous to tables 1 and 8), which suggests that there were some baseline differences in immunisations for DPT, BCG and polio.

Table 9: Pretreatment immunisation

Variables	Treatment group	Control group	Difference
Age-appropriate measles immunisation dose	0.85 (0.008)	0.83 (0.009)	0.011 (0.015)
Age-appropriate DPT immunisation dose	0.77 (0.01)	0.68 (0.01)	0.083** (0.036)
Age-appropriate BCG immunisation dose	0.97 (0.004)	0.94 (0.006)	0.03*** (0.01)
Age-appropriate polio immunisation dose	0.84 (0.008)	0.77 (0.009)	0.065** (0.026)
Vitamin A supplementation (number)	1.0 (0.015)	0.95 (0.015)	0.04 (0.044)

Note: Treatment and control averages are in the middle two columns. Robust standard errors, clustered at facility level, are in parentheses in the difference column. Description of variables: each of the first four outcomes is an indicator for whether the child has received the appropriate dose, given the age. Determination of appropriate immunisations follows the schedule used by the original authors (measles: 1 or 2 for children <2 months; polio: 1, 2 or 3 for children <3 months, 3 or 4 for children >2 months; DPT: 1 or 2 for children <3 months, 3 for children >2 months; BCG: 1 for all children). The fifth outcome is a measure of the number of times the child received a vitamin A supplementation. *Significant at 10%, ** Significant at 5%, *** Significant at 1%.

Since it appears that there may have been some baseline differences in immunisation rates favouring the treatment group prior to the intervention, we then exactly replicate Table 4 from the paper, which estimates the programme's impact on immunisations, but instead use the baseline immunisation data rather than endline data. The results, presented in Table 10, look very similar to the results in Table 4. This suggests that the treatment group had superior immunisation rates before the intervention and maintained this superiority during the study.

Table 10: Impact of treatment assignment on immunisation at baseline

Group	Newborn	Under 1 year	1 year old	2 years old	3 years old	4 years old
Specification:	(1)	(2)	(3)	(4)	(5)	(6)
Average standardised effect	0.12 (0.09)	0.11** (0.055)	0.14*** (0.046)	0.12*** (0.043)	0.11** (0.045)	0.12 (0.079)
Observations	127	827	875	871	893	315

Note: Average standardised effects are derived from equation (3), with the dependent variables being indicator variables for whether the child has received the required dose(s) of measles, DPT, BCG and polio vaccines and vitamin A supplementation, respectively, and including district fixed effects and baseline covariates listed in Table 2. Robust standard errors clustered by catchment areas in parentheses. Groups: (1) children under 3 months; (2) children 0–12 months; (3) children 13–24 months; (4) children 25–36 months; (5) children 37–48 months; (6) children 49–60 months. *Significant at 10%. **Significant at 5%. ***Significant at 1%.

Finally, we explore this result using a difference-in-differences framework, switching our analysis from the individual child level to the facility level, since the outcome is based on binary, permanent indicators (immunisations). In other words, we collapse immunisation rates to the facility level, indicating the percentage of children in each facility's catchment area who had received the appropriate dosage of each immunisation (e.g. per cent of

children receiving age-appropriate measles immunisations). We focus here on children who are newly eligible for all immunisations (those born since the intervention), in order to allow a robust comparison (though the results are not sensitive to which age groups we consider).

We first present the results visually in figures 2 and 3, which show facility-level immunisation rates for newborns (under 3 months) and children under 1 year at baseline and endline. Control group facilities are in solid, darker lines, and treatment group facilities have the lighter dashed lines. Visually, there is not a consistent, clear differential increase in facility-level immunisation rates, and in some cases it appears that the control group facility rates increased by more than those in the treatment group facilities.

Figure 1: Newborn immunisations by facility

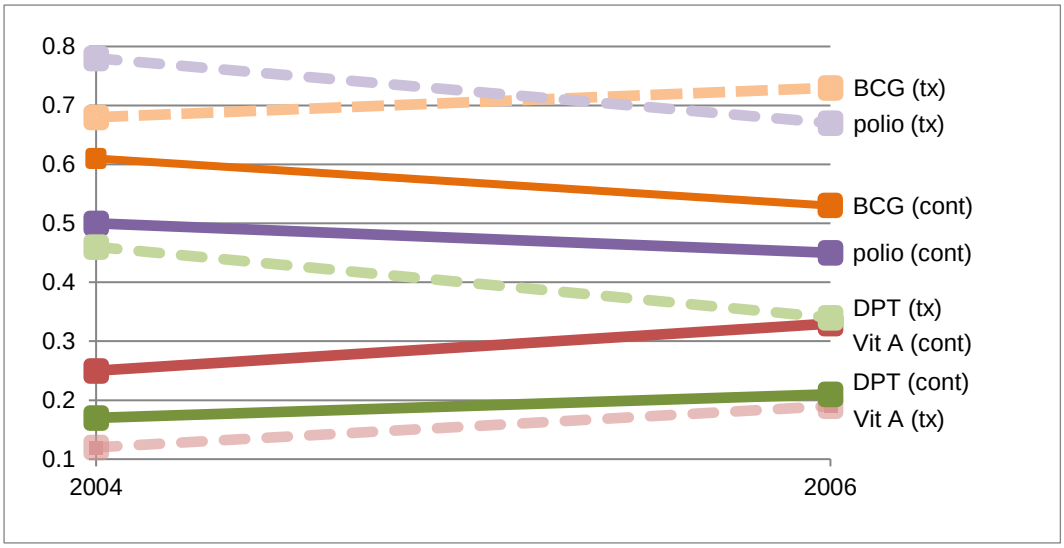
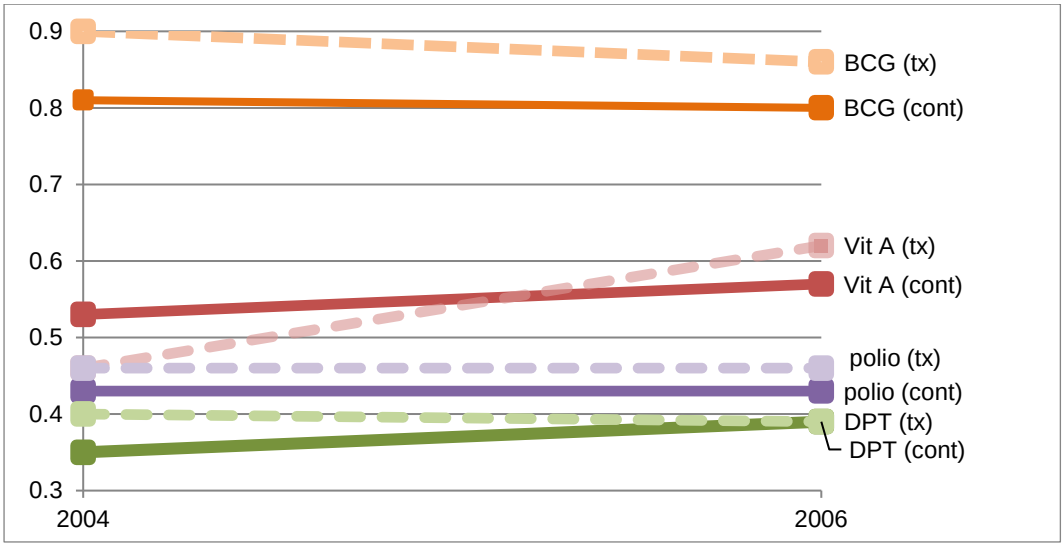


Figure 2: Under-1 immunisations by facility



The analytical results are consistent with the visual outcomes in figures 2 and 3. In Table 11, we present the difference-in-differences analysis for newborns and children under 1 year, analysed at the facility level (and in general following an analysis strategy identical to the

pure replication). While not statistically significant, the results suggest that the programme impact was very small, with no measureable effect.

Table 11: Programme impact on immunisation

Group	Newborn	Under 1 year
Specification	(1)	(2)
Programme impact	0.14 (0.17)	–0.040 (0.13)
Observations	95	100

Note: Programme impact estimates from difference-in-differences models with robust standard errors are reported. Analyses are performed at facility level (per cent of children with appropriate immunisations) and based on standardised effects on DPT, BCG and polio immunisations, in addition to vitamin A supplementation.

4.3 Presence of an implementing CBO prior to the study

The 18 CBOs that implemented the CBM intervention were active in 64 per cent of treatment communities and 48 per cent of control communities prior to the intervention. The authors stated that participating CBOs were focused on health issues, along with some other areas (e.g. 'agricultural development, women's empowerment, support of orphans and vulnerable children, and peace-building initiatives'), and that there were other non-study-related CBOs working concurrently in the study communities. Given that (1) the CBOs' presence could plausibly be influencing or correlated with the communities' health outcomes,[5] and (2) many of the health outcomes are measured only at endline, we believe that it is reasonable to control for whether the CBOs involved in the study were active in the community prior to the intervention (in line with the analyses presented in the last two columns of Table 7[6]).[7] Our pre-analysis plan pre-specified considering specific CBO characteristics instead (e.g. whether CBO leadership is local or not), but the authors were unable to share these data due to ongoing analyses by their own team.

[5] For example, it may be that the participating CBOs were chosen because they were higher quality. It could then be that the observed improved health outcomes are due more to these CBOs' presence than the intervention. Or, perhaps implementing CBOs were already in the communities frequently, and naturally informally emphasising the messages from the treatment. On the other hand, it may be that the CBOs were drawn to relatively worse-off communities, which would suggest that the analysis underreports the impact of the programme.

[6] The specifications in Table 12 differ from those in Table 7 in that they are standard, individual OLS regressions rather than seemingly unrelated regressions, and include the standard set of controls (baseline covariates) from in the main results tables (including Table 6, on which this table expands).

[7] As the authors pointed out to us upon review of an earlier draft, the CBOs' characteristics should be orthogonal to treatment, given random assignment. However, we believe that it is still reasonable to include this control – and that it is, in fact, a standard procedure to include covariates for baseline characteristics that are unbalanced by chance. We argue that this is especially important given that many important outcomes are measured only at endline. The authors also pointed out to us that the interpretation of the coefficients on CBO prior presence is not straightforward, since members of the implementing CBOs received training as part of the intervention. While this is a reasonable assertion, and we agree that the coefficients should be interpreted with caution, the authors included CBOs as a control variable in the main paper, in Table 7, suggesting that this problem is not sufficient to exclude this analysis at all.

We focus our analysis of the impact of incorporating a control for prior CBO presence and an interaction between prior CBO presence and treatment assignment on the results presented in Table 6 (Programme impact on health outcomes), while maintaining everything else exactly as the authors specified. We present these results in Table 12. The results in the first two columns, number of births, appear mostly unchanged as a result of including a control for an intervention CBO's presence prior to the intervention and the interaction with treatment status. The coefficient on number of pregnancies in the household is not affected by including a control for prior CBO presence alone (column 3), but it loses magnitude and significance when additionally including the CBO presence-treatment interaction (column 4).

The results on under-5 mortality, in columns 5 and 6 of Table 12, are moderately different from the originally reported results. By adding a control for whether a participating CBO was present prior to the intervention, the point estimate on the reduction in mortality falls about 15 per cent and loses statistical significance (p=0.2). Including the interaction term leads to a further reduction in magnitude and significance (p=0.59). The coefficients on CBO presence are also imprecise, but also negative, suggesting that the CBOs may have separately helped reduce the mortality rate. The results in column 7, breaking down the child death impacts by year of birth, remain fairly consistent with the original results, apart from some reductions in statistical significance. It continues to appear that the programme was most beneficial in reducing mortality for the youngest children.

The last four columns in Table 12, corresponding to weight-for-age z-scores, similarly show a reduction in magnitude (by 21–29 per cent) and significance (p=0.20, p=0.097) of the coefficient on programme impact when controlling for prior CBO presence but an overall favourable impact of the programme. In these specifications, however, the CBOs' prior presence seems to have an especially beneficial effect on health outcomes and temper the measured treatment effect. Understanding the mechanisms behind this result would likely require more information about the specific nature of the CBOs' work.

Given the results reported above on immunisations (that it is likely that immunisation rates did not measurably improve as a result of the programme), we additionally consider how prior CBO presence influenced baseline immunisation rates. We present the results in table A1 of the appendix and suggest that prior CBO presence is not the primary driver of higher immunisation rates at baseline.

Table 12: Programme impact on health outcomes, controlling for prior CBO presence

Panel A

Dependent variable	Births		Pregnancies		Under-5 mortality rate		Child death
Specification:	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Programme impact	–0.018 (0.015)	–0.011 (0.025)	–0.03** (0.014)	–0.018 (0.021)	–42.6 (32.5)	–33.5 (60.74)	
CBO present prior to intervention	0.012 (0.015)	0.018 (0.022)	0.007 (0.018)	0.018 (0.023)	–34.84 (42.35)	–27.5 (45.0)	–0.009 (0.009)
Programme impact x CBO present prior to intervention		–0.013 (0.034)		–0.025 (0.031)		–15.89 (70.58)	
Child age (log)							
Female							
Programme impact x year of birth 2005							–0.025* (0.013)
Programme impact x year of birth 2004							–0.018* (0.010)
Programme impact x year of birth 2003							0.005 (0.010)
Programme impact x year of birth 2002							0.002 (0.006)
Programme impact x year of birth 2001							0.003 (0.006)
F test on programme impact and CBO present prior to intervention	0.8 –0.46		2.49 –0.093		4.19 –0.024		
F test on CBO present prior to intervention and CBO present x programme impact		0.33 (0.72)		0.38 (0.69)			0.32 (0.73)
F test on programme impact, CBO present prior to intervention, and CBO present x programme impact		0.66 (0.58)		1.77 (0.17)			2.75 (0.059)
Mean control group 2005	0.22	0.22	0.29	0.29	144.4	144.4	0.029
Observations	4,996	4,996	4,996	4,996	50	50	5,094

Table 12: Programme impact on health outcomes, controlling for prior CBO presence

Panel B

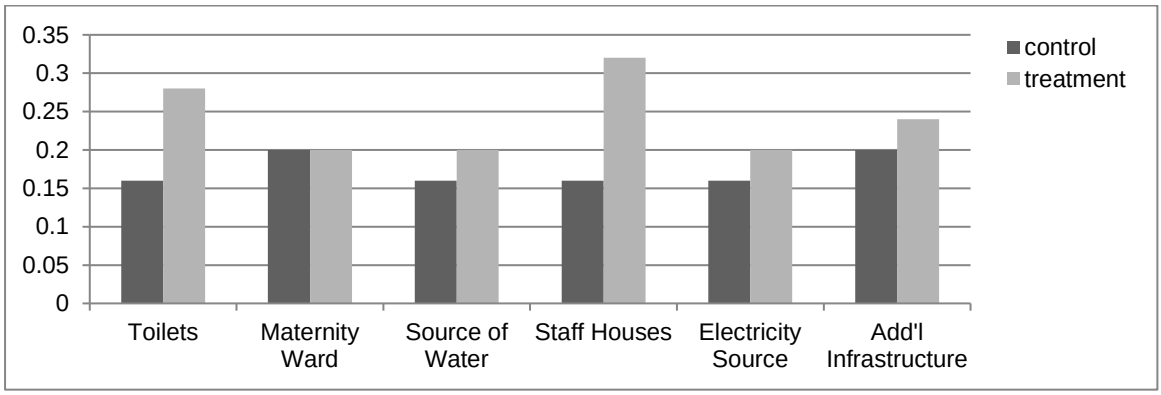
Dependent variable	Weight-for-age z-score			
Specification:	(8)	(9)	(10)	(11)
Programme impact	0.10 (0.07)	0.15 (0.13)	0.11* (0.06)	0.14 (0.11)
CBO present prior to intervention	0.31*** (0.10)	0.36** (0.14)	0.24*** (0.09)	0.27** (0.12)
Programme impact x CBO present prior to intervention		–0.10 (0.15)		–0.065 (0.13)
Child age (log)			–1.27*** (0.07)	–1.27*** (0.07)
Female			0.27*** (0.09)	0.27*** (0.09)
F test on programme impact and CBO present prior to intervention	6.57 –0.003		2.55 –0.088	
F test on CBO present prior to intervention and CBO present x programme impact		4.61 (0.015)		3.81 (0.029)
F test on programme impact, CBO present prior to intervention, and CBO present x programme impact		5.06 (0.0039)		3.77 (0.016)
Mean control group 2005	–0.71	–0.71	–0.71	–0.71
Observations	1,135	1,135	1,135	1,135

Note: Table is exactly as in Table 6 of main text, except that an indicator for whether the study CBO was present in the community prior to the intervention is included in all specifications and an interaction of treatment with prior CBO presence is included where indicated. Estimates from equation (1) with district fixed effects and baseline covariates as listed in table 2 are included. Specification 7 includes a full set of year-of-birth indicators. Robust standard errors in parentheses (3), clustered by catchment area 1–4, 7–11. F statistics and associated p-values are reported where appropriate. *Significant at 10% level. **Significant at 5% level. ***Significant at 1% level.

4.4 Process measures: effort, technology/infrastructure, other

The intervention's policy usefulness is somewhat diminished by the 'black box' nature of going from the intervention to improvements in health (Marston *et al.* 2013). As a result, we and the authors have attempted to gain a better sense of what processes might have changed. The authors present a number of these throughout the original paper, and we verify them in our pure replication. There are, however, a few domains that the authors did not thoroughly explore, which we hypothesise might also be relevant. For example, it could be that treatment health facilities improved their physical infrastructure and capacities as a result of the intervention. There are no statistically significant changes in various measures of infrastructure (not shown), though visually in figure 3 there appear to be some small changes. In no case are treatment facilities less likely to have renovated or built new infrastructure, and in some cases, such as toilets or staff houses, they appear more likely to have made improvements.

Figure 3: Has health facility renovated or built new infrastructure?



Note: All responses come from health facility-level surveys during the post-intervention survey. Respondents were asked whether the health facility had built additional units or made renovations to the toilets, maternity ward, water source, staff houses, electricity source or other infrastructure since January 2005. Mean rates for positive responses to new construction or renovation are shown for treatment and control facilities.

We are similarly interested in understanding to what extent provider effort was influenced by the intervention. We consider two (imperfect) proxies for provider effort, with results presented in Table 13. The first, the number of outreach visits per month, appears to have increased 26 per cent in the treatment group. Using the same difference-in-differences specification seen in Table 3, we estimated a programme impact of 0.89 (p=0.09) additional outreach visits, compared to a control group mean of 4.53 visits. We additionally consider households' awareness of the intervention, finding that households in treatment communities were 40 per cent more likely to know about the intervention. Considering that the treatment communities cover 55,000 households, this is not a trivial effect, and likely at least partially driven by health providers' outreach efforts and consultations with community members. We interpret these effects as reflective of significantly more effort by providers in the treatment facilities, and evidence that the programme would likely scale well.

Table 13: Programme impact on proxies for provider effort

	Number of outreach visits per month	Household is aware of programme
Specification:	(1)	(2)
Programme impact	0.88* (0.51)	0.045*** (0.013)
Observations	96	4,996
Mean control group 2005	4.53	0.11

Note: Programme impact estimates from difference-in-differences models with robust standard errors are reported in column 1, as in Table 5. Column 2 reports the coefficient from a cross-sectional analysis with district fixed effects and baseline covariates as listed in Table 2, with robust standard errors in parentheses.

5. Limitations

Several interesting dimensions of our pre-specified replication extension relied on having more specific information about the report cards that communities received, the action plans that communities developed, the results of the local council survey, and the CBOs that implemented the intervention. Because the authors are writing follow-up papers using this information, they were not able to share these resources with us, but some valuable interesting analyses remain. For example, it would be useful to understand if there is

differential response among providers and/or communities to being given different relative performance rankings, since it seems likely that mechanisms for improvements would be different, depending on how well a facility is reported to perform. Additionally, communities were encouraged to identify priorities for improvement in their action plans, and it would be useful to understand whether communities' improvements match these priorities, how suggested timelines influenced outcomes, how the communities planned to monitor changes and so on. Finally, it would be useful to examine the influence of an implementing CBO's prior presence, particularly why it appears to have had such a strong influence on child weight-for-age z-scores.

After seeing the instruments used to collect the household-level data, it became apparent to us that the enumerators only asked the relevant questions used to form the under-5 mortality measure if the household had at least one *living* child under 5. In other words, if the household had a young child under 5 who had died in the last year and no other children under 5, we would not pick up this death in the data. At endline, 42 per cent of households did not have a living child under 5 years. Although the majority of these households surely did not have a child under 5 who had died in the last year, and there is no obvious reason to suspect that there would have been differential impact across study groups, the potential for measurement error is concerning. Like the authors, we are not able to improve on this measure, given this limitation in the data. We were able to exactly replicate this reduction in under 5 mortality using the same method as the original authors, but we feel that the inability to know whether the 42 per cent of households without a currently living child under 5 had a child die in the previous year is an unfortunate limitation of the data.

Given the need to ask detailed information about healthcare utilisation and health outcomes during the household surveys, there was limited time to ask about household characteristics. As a result, we have limited demographic information, which might be helpful for investigating treatment effects in the future. For example, much of the relevant information is written, and it would be useful to have a better understanding of literacy rates in the community, particularly among caregivers for young children. Finally, it is not clear to what extent the results from this study can be generalised to the rest of Uganda, or more broadly to other developing countries. This study was conducted in only nine districts of Uganda, and it is not clear whether there were any characteristics that influenced the effect of the intervention.

6. Discussion

In our pure replication of the authors' published results, we have been able to closely verify all of the results, with a caveat on the standardised treatment effects discussed in section 3.1. This suggests that the CBM programme increased monitoring, improved treatment practices and management, improved immunisation rates, improved utilisation rates, reduced under-5 mortality and improved children's weight-for-age z-scores. Verification of the authors' originally presented findings, under the assumptions in the paper, is important for policymakers considering similar programmes.

Our extensions of the paper give more credence to some aspects of the study and raised some important doubts. We show that household characteristics across study arms were balanced, helping alleviate concerns that some differences – particularly the health

outcomes that were measured only at endline – might have been due to pre-existing wealth differences or other household characteristics.

On the other hand, our analysis of immunisation rates at baseline raises important questions about the impacts shown on immunisations. The evidence is consistent with a situation in which treatment communities did not actually improve immunisation rates, since the same 'treatment effect' was already present at baseline. The implicit measurement error related to the measurement of under-5 mortality, combined with the coefficient on under-5 mortality losing magnitude and significance when adding a control for preintervention CBO presence, similarly casts some uncertainty on this important result in the paper. The results presented here certainly do not negate the finding, but they do suggest that the under-5 mortality result should be interpreted with caution. Finally, the finding that the effects on weight-for-age z-scores decline and lose significance when the prior presence of an implementing CBO is included as a control is further reason for pause when considering the health impacts of the intervention.

The authors' examination of process measures that may have changed, combined with our short extension of this exercise, suggest that provider behaviour did improve as a result of the CBM intervention. Regardless of interpretation on the health outcomes, this is an important result that should not be overlooked. Although improving health outcomes is an important goal, the intervention did, after all, target provider behaviour. It was clearly successful at this first-stage goal and expanding on this approach in contexts similar to Uganda's to gain a better understanding of the ideal implementation approach appears warranted.

7. Conclusions

Through this replication exercise, we have been able to verify all of the authors' results, based on the specifications laid out in the original paper. Our extension to these analyses further boosts the findings related to provider behaviour and process measures that may have changed as a result of the programme. There were clearly improvements in provider behaviour, generally without being accompanied by increased outside funding or support from superiors in the public healthcare system. Given that CBM is, most fundamentally, targeted at improving provider effort, the programme was successful.

Our extension analyses focused on health outcomes suggest that policymakers should be more cautious in interpreting the results around immunisation, under-5 mortality and weight-for-age z-scores. Although the evidence we present surely would not rule out policy-relevant improvements in health outcomes, we do identify some important inconsistencies in this story that raise uncertainty about the magnitude of these effects.

Appendix

Figure A1: Respondent Education Level

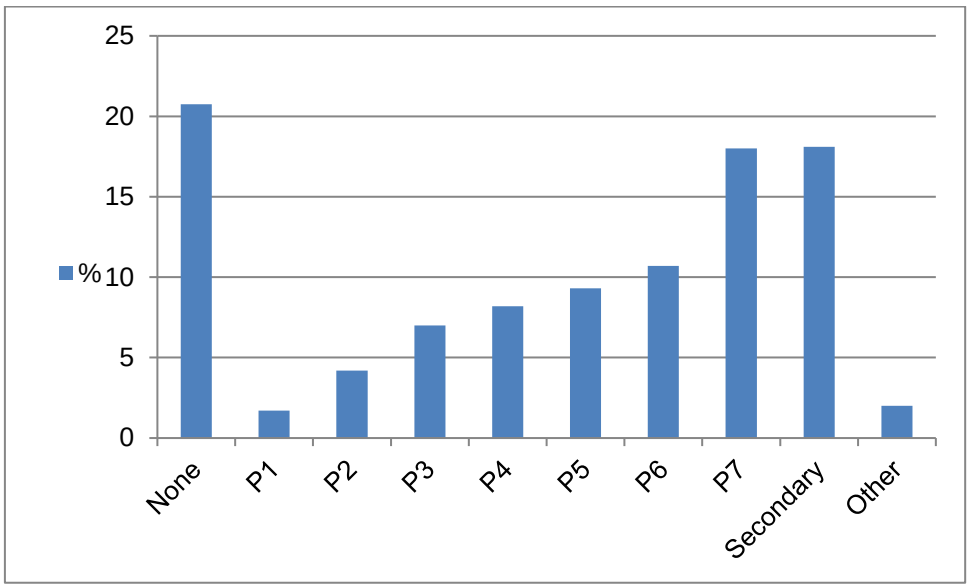


Table A1: Program impact on immunisation, for all cohorts

Group	Newborn	Under 1 year	1 year old	2 years old	3 years old	4 years old
Specification	(1)	(2)	(3)	(4)	(5)	(6)
Programme Impact	0.14	-0.040	-0.22*	-0.60***	-0.12	-0.15
	(0.17)	(0.13)	(0.13)	(0.20)	(0.19)	(0.19)
Observations	95	100	100	100	100	98

Note: Program impact estimates from difference-in-differences models with robust standard errors are reported. Analyses are performed at the facility-level (i.e., per cent of children with appropriate immunisations), and are based on standardised effects on DPT, BCG, and polio immunisations, in addition to vitamin A supplementation.

Table A2: Impact of treatment assignment and prior CBO presence on immunisation at baseline

Group	Newborn	Under 1 year	1 year old	2 years old	3 years old	4 years old
Specification:	(1)	(2)	(3)	(4)	(5)	(6)
Programme impact	0.17*	0.11*	0.15***	0.11**	0.11**	0.11
	(0.10)	(0.056)	(0.056)	(0.048)	(0.044)	(0.076)
CBO present prior to intervention	-0.22*	0.001	-0.019	0.052	0.029	0.20**
	(0.13)	(0.073)	(0.074)	(0.071)	(0.041)	(0.084)
Observations	127	827	875	871	893	315

Note: Average standardised effects are derived from equation (3) with the dependent variables being indicator variables for whether the child has received the required dose(s) of measles, DPT, BCG and polio vaccines and vitamin A supplement, respectively, and with district fixed effects and baseline covariates listed in Table 2 included. Robust standard errors clustered by catchment areas in parentheses. Indicators for whether an implementing CBO was present in the community prior to the project are included in all specifications. Groups: (1) Children under 3 months; (2) Children 0–12 months; (3) Children 13–24 months; (4) Children 25–36 months; (5) Children 37–48 months; (6) Children 49–60 months.

*Significant at 10%, **Significant at 5%, ***Significant at 1%.

References

Banerjee, AV, Banerji, R, Duflo, E, Glennerster, R and Khemani, S, 2010. Pitfalls of participatory programs: evidence from a randomized evaluation in education in India. *American Economic Journal: Economic Policy*, 2(1), pp. 1–30.

Banerjee, A, Deaton, A and Duflo, E, 2004. Health, health care, and economic development: wealth, health, and health services in rural Rajasthan. *The American Economic Review*, 94(2), p. 326.

Barr, A, Mugisha F, Serneels P and Zeitlin A, 2012. Information and collective action in community-based monitoring of schools: field and lab experimental evidence from Uganda. *Unpublished paper*.

Björkman, M and Svensson, J, 2009. Power to the people: evidence from a randomized field experiment on community-based monitoring in Uganda. *The Quarterly Journal of Economics*, 124(2), pp. 735–69.

Chaudhury, N, Hammer, J, Kremer, M, Muralidharan, K and Rogers, FH, 2006. Missing in action: teacher and health worker absence in developing countries. *The Journal of Economic Perspectives*, 20(1), pp. 91–116.

Cohen, J and Dupas, P, 2010. Free distribution or cost-sharing? Evidence from a randomized malaria prevention experiment. *Quarterly Journal of Economics*, 125(1), p. 1.

Cortinovis, I, Vella, V, Ndiku, N and Milani, S, 1997. Weight, height and arm circumference of children under 5 in the district of Mbarara, south-west Uganda. *Annals of Human Biology*, 24(6), pp.557-68.

de Laat, J, Kremer, M and Vermeersch, C. Local participation and teacher incentives: evidence from a randomized experiment. Unpublished manuscript. 2008.

Donato, K and Garcia Mosqueira, A, 2013. Technical proposal for replication of ‘Power to the People: Evidence from a Randomized Field Experiment on Community-Based Monitoring in Uganda’ [online]. Available at: <http://www.3ieimpact.org/media/filer_public/2014/05/09/donato_revised_replication_plan.pdf>

Donato, K and Garcia Mosqueira, A, 2016. Replication documents for ‘Power to the people?: a replication study of a community-based monitoring programme in Uganda’ [online]. Available at: <https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi%3A10.7910%2FDVN%2FWYHDX2> doi:10.7910/DVN/WYHDX2, Harvard Dataverse.

Jones, G, Steketee, RW, Black, RE, Bhutta, ZA and Morris, SS, 2003. How many child deaths can we prevent this year? *The Lancet*, 362(9377), pp. 65–71.

Kidane, G and Morrow, RH, 2000. Teaching mothers to provide home treatment of malaria in Tigray, Ethiopia: a randomised trial. *The Lancet*, 356(9229), pp. 550–55.

Kling, JR, Liebman, JB and Katz, LF, 2007. Experimental analysis of neighborhood effects. *Econometrica*, *75*(1), pp.83–119.

Lindelow, M and Serneels, P, 2006. The performance of health workers in Ethiopia: results from qualitative research. *Social Science & Medicine*, 62(9), pp. 2,225–35.

Mansuri, G and Rao, V, 2013. Localizing development: does participation work? *Policy Research Report*. Washington, DC: World Bank. [online] Available at: <https://openknowledge.worldbank.org/handle/10986/11859>

Marston, C, Renedo, A, McGowan, CR and Portela A, 2013. Effects of community participation on improving uptake of skilled care for maternal and newborn health: a systematic review. *PLOS One,* 8.2: e55012.

McPake, B, Asiimwe D, Mwesigye F, Ofumbi M, Ortenblad L, Streefland P and Turinde A, 1999. Informal economic activities of public health workers in Uganda: implications for quality and accessibility of care. *Social Science & Medicine*, 49(7), pp. 849–65.

Miguel, E and Kremer, M, 2004. Worms: identifying impacts on education and health in the presence of treatment externalities. *Econometrica*, 72(1), pp. 159–217.

Olken, BA, 2007. Monitoring corruption: evidence from a field experiment in Indonesia. *Journal of Political Economy*, 115(2).

Reinikka, R and Svensson, J, 2004. Local capture: evidence from a central government transfer program in Uganda. *The Quarterly Journal of Economics*, 119(2), pp. 679–705.

World Bank, 2014. Health. Available from: <http://data.worldbank.org/topic/health>. [14 August 2016]

World Health Organization, 2016. Children: reducing mortality. Available from: <http://www.who.int/mediacentre/factsheets/fs178/en/>. [14 August 2016]

Publications in the 3ie Replication Paper Series

The following papers are available from http://www.3ieimpact.org/en/publications/3ie-replication-paper-series/

Fighting corruption does improve schooling: a replication study of a newspaper campaign in Uganda, 3ie Replication Paper 10. Kuecken, M, and Valfort, MA (2016)

The effects of land titling on the urban poor: a replication of property rights, 3ie Replication Paper 9. Cameron, Drew B, Whitney, Edward M and Winters, Paul C (2015)

Male circumcision and HIV acquisition reinvestigating the evidence from young men in Kisumu, Kenya, 3ie Replication Paper 8. Djimeu, EW, Korte, JE and Calvo, FA (2015)

Walking on solid ground: a replication study on Piso Firme's impact, 3ie Replication Paper 7. Basurto, MP, Burga, R, Toro, JLF and Huaroto, C (2015)

The impact of India's JSY conditional cash transfer programme: A replication study, 3ie Replication Paper 6. Carvalho, N and Rokicki, S (2015)

Recalling extra data: A replication study of finding missing markets, 3ie Replication Paper 5. Wood, BDK and Dong, M (2015)

The long and short of returns to public investments in fifteen Ethiopian villages, 3ie Replication Paper 4. Bowser, WH (2015)

Reanalysis of health and educational impacts of a school-based deworming program in western Kenya Part 2: Alternative analyses, 3ie Replication Paper 3, part 2. Aiken, AM, Davey, C, Hayes, RJ and Hargreaves, JR (2014)

Reanalysis of health and educational impacts of a school-based deworming program in western Kenya Part 1: A pure replication, 3ie Replication Paper 3, part 1. Aiken, AM, Davey, C, Hargreaves, JR and Hayes, RJ (2014)

TV, female empowerment and demographic change in rural India, 3ie Replication Paper 2. Iversen, V and Palmer-Jones, R (2014)

Quality evidence for policymaking: I'll believe it when I see the replication, 3ie Replication Paper 1. Brown, AN, Cameron, DB and Wood, BDK (2014)

Replication Paper Series

International Initiative for Impact Evaluation
1029 Vermont Avenue, NW
Suite 1000
Washington, DC 20005
USA

replication@3ieimpact.org
Tel: +1 202 629 3939

www.3ieimpact.org