Does Performance Evidence Motivate? A Field Experiment in Guinea-Bissau's Health Sector

Mattia Fracchia*

Job Market Paper

This versions: November 6, 2023

[Click here for the most recent version]

Abstract

This paper investigates how employees respond to evidence about their organization's performance in achieving its mission. In a randomized field experiment with the Ministry of Public Health of Guinea-Bissau, I provide front-line health workers with evidence of their program's effectiveness in improving local health indicators. The intervention significantly enhanced healthcare service delivery performance, with treated workers providing 39 percent more healthcare time to recipients compared to the control group six months post-intervention. The intervention is particularly effective in correcting over-optimistic prior beliefs about the status of health indicators. These results highlight the potential for organizations to harness performance evidence to enhance workers' efforts.

JEL codes: D83, D84, D91, I15, J24, O10, O12

Keywords: Public Service, Delivery, Evidence, Motivation, Experiment, Guinea-Bissau, Africa.

*Nova School of Business and Economics, NOVAFRICA. Email: mattia.fracchia@novasbe.pt. I am indebted to Pedro C Vicente for his guidance and support throughout this project. I also thank seminar and conference participants at NEUDC 2023 at Harvard University, World Bank DIME-KDI School 4th Development Impact Conference, University of St.Gallen, Oxford Development Economics Workshop, Advances in Field Experiments at The University of Chicago, SITES at University of Naples Parthenope, UNU-EGOV, Royal Economic Society Ph.D. Conference at The University of Glasgow, Transatlantic Doctoral Conference at London Business School, Yale-RISE Conference, NOVAFRICA Conference on Economic Development. I gratefully acknowledge financial support from the Research Unit Programmatic Funding scheme at Nova SBE, the SurveyCTO Data Collection Research Grant, and the Bissau Economics Lab. I thank the Regional Health Directorates of Farim and Oio, and Radio Sol Mansi for generous support during field activities. Universidade Nova de Lisboa and the National Committee for Ethics in Research in Health of Guinea-Bissau gave ethics clearance. The study was pre-registered at the American Economic Association (AEARCTR-0009054). All errors remain my own.

1 Introduction

Many organizations adopt missions centered around social impact or environmental sustainability. The economics literature has shown that workers care about the mission of their job (Besley and Ghatak, 2005), which proved to improve selection and enhance performance. However, while any organization can claim an ambitious mission, not all will fulfill it with the same success (Fioretti, 2022). How do workers respond to evidence about their organization's performance in achieving the claimed mission?

In this paper, I explore how employees respond to evidence of their organization's performance in achieving its mission and whether an organization may leverage it to motivate workers. These questions are particularly relevant for public service organizations in Low and Middle-Income Countries (LMICs) for two main reasons. First, governments worldwide face challenges in motivating workers in public service delivery (Finan, Olken, and Pande, 2017). Oftentimes, conventional tools standard economics advocates to motivate workers, such as financial incentives, are unavailable or unfeasible. Second, as donors and policymakers increasingly recognized the value of evidence-based decision-making in recent years, available evidence grew in scope, rigor, and accessibility. Much of its potential is still untapped (World Bank, 2021).

The effects of organizational communication shifting from unverified mission claims toward performance evidence are a priori uncertain. On the one hand, workers may value the status quo of unverified, well-intentioned claims as they meet their psychological needs, such as the fundamental desire to preserve a positive self-image (Steele, 1988). On the other hand, the lack of evidence may lead workers to perceive the mission claims as mere rhetoric, ultimately leading to detachment. When exposed to new evidence, workers may face a trade-off between belief desirability and accuracy (Bénabou and Tirole, 2016). As they value the mission of their job, they may want to protect self-serving beliefs by ignoring the new evidence ex-ante (Dana, Weber, and Kuang, 2007; Exley and Kessler, 2021; Golman, Hagmann, and Loewenstein, 2017; Grossman and Van Der Weele, 2017) or distort its interpretation ex-post (Eil and Rao, 2011; Möbius et al., 2022; Zimmermann, 2020).[1] Moreover, workers could interpret new evidence as altering the marginal return to mission-related effort[2] or shifting the

[1] Ex-ante avoidance of information and ex-post distortion of beliefs can explain phenomena like managerial overconfidence (Huffman, Raymond, and Shvets, 2022; Malmendier and Tate, 2005) and can spread quickly through organizations (Bénabou, 2013).

[2] An influential literature studied how perceived future returns influence investments in education (Jensen, 2010; Kaufmann, 2014; Nguyen, 2008).

perceived distance from the target goal (Kőszegi and Rabin, 2006; Abeler et al., 2011).[3]

To study workers' responses to communications of performance evidence and analyze its causal impact on effort, I design a field experiment in collaboration with a nationwide public health program in Guinea-Bissau. The program has the only claimed mission of improving maternal, newborn, and child health. In the experiment, front-line service providers, representing the program's backbone, are randomly assigned to receive correct and up-to-date evidence of the program's performance in achieving the claimed mission. The evidence includes data on the program's target health indicators at the local and national levels, their evolution before and after the program's introduction, and the program's effectiveness in achieving its pre-defined goals. I measure how the treatment influences workers' beliefs and efforts using multiple surveys of workers and the program's recipients.

I design and implement an evidence-provision intervention in partnership with the Community Health Workers (henceforth, CHWs) program of the Ministry of Public Health of Guinea-Bissau. CHWs are community members who perform door-to-door household visits to link their communities to the health system while directly providing selected health services. However, the program is unable to monitor either the visits or the associated activities, potentially leading workers to shirk their responsibilities. Hence, the centrality of keeping CHWs motivated and engaged. Under treatment, workers receive a module combining *anecdotal* and *scientific* evidence on their program's performance in achieving its mission. The evidence focuses on key health indicators at the local level before and after the CHWs program introduction, showing how they score compared with the program goals and how the program contributed to the results shown. Workers are encouraged to share thoughts about the module's content and their experiences relating to the program's mission and impact. The control group receives a placebo module to keep the workshop structure comparable to the treatment group. The treatment and the placebo modules are followed by skill-refresher training for the workers.

The main outcome of interest is healthcare service delivery through door-to-door household visits, the key indicator of CHWs' performance. I conduct a phone survey with a sample of three to four households per CHW six months after the intervention. Households accompanied by a CHW assigned to the treatment group report receiving 11 percent more visits the month before the interview

[3] Thakral and Tô (2021) provide recent field evidence on how reference points adjust. This represents the latest contribution to the debate sparked by the evidence of negative wage elasticities among New York City taxicab drivers in Camerer et al. (1997). Over the year, this debate saw the contributions of Chou (2002), Crawford and Meng (2011), Doran (2014), and Farber (2005, 2008, 2015).

(November 2022), which lasted, on average, 22 percent longer than the control group. This translates into 39 percent more healthcare service time delivered by workers in the treatment group relative to the control group. The positive treatment effect on healthcare service delivery is present, although less precisely estimated, for the six months covering the period from the intervention to the household survey. Furthermore, I collect a set of measures of the quality of the service delivered. Households accompanied by a CHW assigned to the treatment group receive more healthcare practices promotion, which is the CHW's central task during the door-to-door household visits, compared to those accompanied by a CHW in the control group. Finally, households declare being 3 percent more satisfied with the program when accompanied by a CHW in the treatment group.

The increase in healthcare services received by the program's recipients is consistent with CHWs' self-reported intentions. I conduct a phone survey with CHWs four months after the intervention and elicit CHWs' intentions to deliver healthcare services to households in the next month (November 2022). CHWs assigned to the treatment group report the intention to spend 27 percent more time with households during the next month. The increase in willingness to exert effort in the program emerges already during the workshop immediately after the intervention. CHWs assigned to the treatment group report planning to work 14 percent more time in the week after the intervention. Moreover, CHWs in the treatment group appear to learn more during the skill-refresher training following the intervention, showing an overall 6 percent larger grade improvement between pre and post-training tests than the control group.

To study how CHWs respond to evidence about their organization's performance in achieving the claimed mission, I elicit two sets of beliefs on six program's key target health indicators. The intervention directly addresses three of them. First, for all CHWs, I elicit prior and posterior beliefs about the status of the six health indicators at the local level. On average, workers assigned to the treatment group improved the accuracy of their beliefs about the six program's target health indicators by 20 percent relative to the control group. The intervention was particularly effective at correcting the most inaccurate and over-optimistic beliefs. In fact, the improvement in accuracy comes entirely from a downward revision of prior beliefs. Second, for each health indicator, I elicit CHWs' prior and posterior perceptions of the program's effectiveness in improving it. I find that the intervention reduced the average perceived program's effectiveness in improving health indicators at the local level by 3 percent. The intervention was particularly effective at correcting the more optimistic beliefs

about the program's effectiveness.

It is worth noting that the intervention relies on existing evidence of the organization's effectiveness in achieving its mission. In contexts with limited resources, organizations might be unable to supervise workers closely and provide tailored feedback. Hence, leveraging existing evidence at the organizational level may result in a cost-effective tool to achieve performance enhancement. The intervention is, in fact, highly cost-effective (Appendix Section K). I estimate that providing the experimental evidence-sharing module costs 2.91 USD per CHW. As a result, in the setting of this study, increasing healthcare time delivered by 1 standard deviation has a per capita cost of 5.56 USD. These results confirm the feasibility of leveraging existing evidence of mission accomplishments to enhance workers' performance even in settings with scarce resources.

This paper provides the first experimental evidence that an organization can motivate workers by sharing evidence of the organizational performance in achieving its mission. Thus, it contributes to the vast and diverse literature on incentives in organizations (Ashraf and Bandiera, 2018; Ellingsen and Johannesson, 2008; Kamenica, 2012; Rotemberg, 1994). Theoretically, the economic literature has widely and deeply investigated the multidimensional nature of motivation. This paper shares in common with Akerlof and Kranton (2000, 2005), Bénabou and Tirole (2003, 2006), Delfgaauw and Dur (2008), Dixit (2002), Francois (2000), and Murdock (2002) the focus on the non-pecuniary dimensions of motivation. In particular, it adds to the notion that workers get utility from contributing to the organization's mission and that managers can leverage it (Besley and Ghatak, 2005; Cassar and Meier, 2018; Cassar and Armouti-Hansen, 2020; Khan, 2020).[4] In this sense, this paper contributes to the existing literature on management practices in both private firms (Bloom and Van Reenen, 2010; Bloom et al., 2013) and public organizations (Bloom et al., 2015; Fenizia, 2022; Rasul and Rogger, 2018).[5]

This paper contributes directly to the literature on improving state capacity and public service

[4]It is important to note that monetary rewards for promoting impact-related outcomes present several challenges. The impact can be multifaceted and hard to measure. Moreover, incentives directly targeting extrinsic motivation may backfire by crowding out the prosocial value motivated workers derive from achieving the mission (Frey and Jegen, 2001; Gneezy, Meier, and Rey-Biel, 2011).

[5]By exploring workers' response to a shift from well-intentioned claims toward tangible mission accomplishments, this paper speaks to the literature on Corporate Social Responsibility (CSR) (Bénabou and Tirole, 2010) and Environmental, Social, and Governance (ESG) (Gillan, Koch, and Starks, 2021). This literature has witnessed a gradual shift in focus from measuring activities to impact assessment (Barnett, Henriques, and Husted, 2020). This contribution may be particularly relevant for organizations adopting mission-driven models with social or environmental impact as their primary objective (Besley and Ghatak, 2018; Gertner, 2023).

delivery in countries with weak institutions (Besley et al., 2022). Previous studies in contexts with limited contract enforcement have focused on enhancing selection (Ashraf et al., 2020; Dal Bó, Finan, and Rossi, 2013; Deserranno, 2019) or designing incentives to improve workers' performance (Ashraf, Bandiera, and Lee, 2014; Ashraf, Bandiera, and Jack, 2014; Bandiera et al., 2023; Björkman Nyqvist et al., 2019; Chaudhury et al., 2006; Deserranno, Kastrau, and León Ciliotta, 2021; Deserranno et al., 2020; Duflo, Hanna, and Ryan, 2012; Giné, Mansuri, and Shrestha, 2022; Glewwe, Ilias, and Kremer, 2010; Khan, Khwaja, and Olken, 2016). This paper adds to this literature by showing that organizations can leverage existing evidence of mission accomplishment to enhance performance without altering incentives. Moreover, using rigorous empirical evidence, this paper goes beyond previous studies that sought to elicit motivation through interventions leveraging mission motivation (Khan, 2020) and task significance (Fracchia, Molina-Millán, and Vicente, 2023).[6]

Finally, the intervention of this paper relates directly to the recent literature studying how empirical evidence of policy effectiveness influences the beliefs formation and decision-making of politicians (Hjort et al., 2021), policymakers (Dunning et al., 2019a; Vivalt and Coville, 2023), policy professionals (Banuri, Dercon, and Gauri, 2019), and bureaucrats (Rogger and Somani, 2023). However, this information seldom reaches people on the ground, who may hold inaccurate beliefs about the status of the program's target indicators and its effectiveness in improving them. This paper adds to this literature by targeting workers instead of decision-makers and measuring performance as an outcome.[7]

The paper is organized as follows. In the next section, I give the context of the experiment. Section 3 is dedicated to experimental design, describing treatments, sampling and randomization, measurement, and balance. The following section explains the hypotheses and estimation strategy. Subsequently, I show the econometric results. Section 6 discusses cost-effectiveness and scaling up. Section 7 concludes.

[6] A complementary literature shows that donors tend not to react to evidence of program effectiveness (Metzger and Günther, 2019b) and charitable donations appear unresponsive to the social benefit of giving (Karlan and Wood, 2017; Metzger and Günther, 2019a; Null, 2011). On the contrary, donors may use charity performance metrics as an excuse not to give (Exley, 2020).

[7] It is beyond the scope of this paper to study workers' demand for evidence, which is a priori uncertain. On the one hand, workers may value the status quo of unverified well-intentioned claims as they meet their psychological needs, such as the fundamental desire to preserve a positive self-image (Steele, 1988). On the other hand, the lack of accuracy may lead workers to perceive the mission claims as mere rhetoric, ultimately leading to detachment.

2 Context

I design and implement the field experiment in partnership with the CHWs program of the Ministry of Public Health of Guinea-Bissau. In Guinea-Bissau, as in many LMICs, rural populations frequently lack access to essential public services. More than two-thirds of the population lives below the poverty line.[8] Its GDP per capita in current USD was 727.5 in 2020, ranking 247 in 261 countries (World Bank, 2020). The country's life expectancy in 2021 was 59.65 years, lower than the average of 60.24 in Sub-Saharan Africa in the same year (World Bank, 2023). The country persists in high maternal and infant morbidity and mortality and lags behind in the path to achieving Sustainable Development Goal 3, which aims to "*ensure healthy lives and promote well-being for all at all ages*". Guinea-Bissau scored 8th globally for Neonatal Mortality Rate with 34.30 deaths per 1,000 live births (2021), 9th for Maternal Mortality Rate with 725 deaths per 100,000 live births (2020), and 20th for Under 5 Mortality Rate with 74.30 deaths per 1,000 live births (2021) (World Bank, 2023). The national health system faces persistent challenges related to inadequate health workers, low public spending, and poor infrastructure (World Bank, 2019).[9]

To improve populations' access to primary healthcare while ensuring financial sustainability, governments worldwide have adopted CHWs programs, which involve recruiting individuals from within the villages and providing them with training to serve their communities. Over the past decades, these programs helped prevent and treat prevalent health problems - see Haines et al. (2007), Christopher et al. (2011) and Gilmore and McAuliffe (2013). While the potential of these programs is widely recognized, their implementation poses numerous challenges. The crucial component to the success of these programs is CHWs' performance. However, effectively monitoring these workers is challenging due to their placement in remote areas. Moreover, due to binding budget constraints, CHWs are often non-salaried and operate with a small monthly allowance.[10] They tend to have flexible working hours and often engage in part-time employment alongside other self-managed activities.[11]

[8]Since independence, Guinea-Bissau has been characterized by a 'stable political instability' (Russo et al., 2017). Several coups and coups attempts undermined its capacity to attract external assistance. The last coup attempt took place on February 2, 2022 - see The Economist.

[9]The latest data available report that Guinea-Bissau had only 0.13 physicians per 1,000 people in 2016, compared to an average of 0.21 in Sub-Saharan Africa, and 0.69 nurses and midwives per 1,000 people in 2018, compared to an average of 1.01 in Sub-Saharan Africa (Our World in Data team, 2023).

[10]Data for 34 Sub-Saharan African countries on 322,199 CHWs show that 69 percent of these workers do not receive a regular salary, and 46 percent do not receive any monetary compensation. Data Tool, One Million Community Health Workers Campaign, 2019.

[11]The path to professionalization has been successful (Zambruni et al., 2017). However, it is not affordable

Designing effective incentive schemes remains a significant challenge.

The Ministry of Public Health of Guinea-Bissau introduced the CHWs program in 2017 with the support of international organizations.[12] The program has the explicit mission of improving maternal, newborn, and child health, following international guidelines. It involves recruiting community members through a local selection process organized by community representatives.[13] Once recruited, all CHWs undergo a two-week training session at the local health center. This training equips them with the knowledge and skills to provide health education, identify and prevent diseases, refer households to health centers, and administer basic medical treatments during household visits.[14] Upon completing the training, CHWs are expected to visit all households within their pre-assigned group each month – typically around 60 households. The position is not salaried, although they receive a monthly allowance to cover transportation and communication costs.[15] Supervision is loose. The supervisors are selected among doctors, nurses, and administrative personnel from the local health center, with limited time and resources to supervise the CHWs. Most of the supervision takes place in the monthly meetings, where CHWs also receive refresher training.

3 Experimental design

Figure B1 in the Appendix presents a timeline of the experiment. The intervention was delivered during a one-day workshop at the health center level, replicating the monthly meetings of the CHWs program. Figure B2 presents the program of the workshop.

in many LMICs without significant inflows of foreign aid (Taylor, Griffiths, and Lilford, 2017).

[12]The program was initially managed by non-governmental organizations (NGOs) and later integrated into the public health system. From 2017 to 2019, the European Union and UNICEF were the program's main funders. NGOs implemented activities, and payments were processed at the health center level. In 2019, funding moved to the World Bank, implementation to the Regional Health Directorates, and payments were processed at the central level via mobile money. The program's activities were disrupted, with long delays in the monthly payments. The consequence was a stop in the program's activities and strikes among CHWs.

[13]The position of a CHW is advertised without any mention of monetary compensation or career prospects.

[14]Appendix Section A presents the essential health practices promoted by CHWs.

[15]The monthly allowance is a function of the number of household visits CHWs perform – approximately USD 0.3 per household visit. Since CHWs are expected to visit around 50 households, the typical monthly allowance is around USD 16. Applying the relevant PPP exchange rate, the value corresponds to USD 46.

3.1 Treatments

The intervention conveys to front-line workers involved in the nationwide CHWs program correct and up-to-date evidence of its effectiveness in achieving its mission.[16] I designed the intervention with a team composed of a General Practitioner and a Pediatrician from the largest hospital in the country[17] and two nurses who held the role of supervisors during the CHWs program's first phase (2017 - 2019), in other rural areas of the country. The two nurses delivered the intervention, leveraging their experience communicating with CHWs in rural areas. All communication materials were developed and delivered in Creole, the *lingua franca* in Guinea-Bissau. I now turn to the details of the intervention.

The intervention combines *anecdotal* and *scientific* evidence. The introductory anecdote serves as a tool to engage participants, enabling them to approach the topic by vividly envisioning how changes in health indicators can manifest in everyday life. It mirrors the chronological framework that will be employed in the subsequent scientific evidence, explicitly linking specific months to well-recognized seasonal events like the cashew campaign, carnival festivities, and the rainy season. The trainer adheres to a predefined script (a summary of which can be found in Appendix Section C.1) to ensure a structured narrative. To facilitate participants' understanding, we employ a series of accompanying slides featuring images and key information. This narrative chronicles the journey of a family over the same time period covered by the scientific evidence module, highlighting the transformative impact of the CHWs program on healthcare within the village. It achieves this by incorporating culturally relevant elements, such as the influential roles of matriarchs and traditional healers as the village's initial healthcare providers, as well as traditional beliefs concerning the potential threats posed to children by malevolent supernatural forces.

The scientific part draws evidence from the official program evaluation report by UNICEF Guinea-Bissau.[18] The intervention truthfully conveys the report's main findings, which focus on the program's target health indicators. All CHWs in the treatment group received the same evidence.[19] Among the program's target indicators, we select i) under-five mortality rate (Figure C2), ii) share of pregnant women completing at least four antenatal checks (Figure C3), and iii) share of women giving

[16] Importantly, policymakers and donors already had access to this evidence at the central level.
[17] Hospital Nacional Simão Mendes in Bissau.
[18] The program evaluation is based on UNICEF's Multiple Indicator Cluster Surveys (MICS).
[19] The indicators' most granular level of disaggregation is the administrative region level. All CHWs in the sample operated in the Oio administrative region.

birth assisted by skilled health personnel (Figure C4).[20] For each indicator, we created infographics to help the trainers explain the evaluators' analysis of the program's contribution to achieving the results shown as follows.

1. **Before and after**: The levels of the target health indicators at the local level in 2014, before the program introduction, and in 2019, at the end of the program's first phase.[21]

2. **Target achievement**: The program evaluation assesses the program's effectiveness by measuring where 2019 health indicators stand compared to the values the program set as explicit goals as part of its mission of improving health at the community level. The intervention reports this comparison and conveys the analysis carried out in the program evaluation report.

3. **Local vis-à-vis National**: The program set its goals based on 2019 national-level health indicators. However, there is significant heterogeneity in the status of health indicators across the country's administrative regions. The intervention includes the value of health indicators at both the local and national levels to allow for a comparison.

The UNICEF evaluation, as is often the case with large-scale public delivery programs in LMICs, did not assess the causal impact of the program. Hence, the intervention provides a disclaimer about the correlational nature of the analysis. It complements it with the causally identified results on under 5 mortality reduction of a CHWs program in Uganda from Björkman Nyqvist et al. (2019), which are qualitatively compatible with those reported in the program evaluation.[22] As part of the intervention, workers were encouraged to actively participate and share thoughts about the intervention's content and their experiences relating to the program's effectiveness in achieving its mission.

The *Control* group received a placebo module to keep the workshop structure comparable to the treatment group. The placebo discusses mangoes, their cultural relevance, and their use in local cuisine. As the workshops took place during mango season in Guinea-Bissau, the placebo leveraged

[20]The program evaluation report did not assess the program effectiveness in reducing maternal mortality as MICS did not collect data on this indicator for 2019.

[21]For the under-five mortality rate we provided national-level indicators due to data availability.

[22]Impact evaluations present the clear advantage of providing causal evidence. They, however, present potential caveats which may backfire with front-line service providers. First, publication bias may lead to exaggerating the effectiveness of policies (Casey, Glennerster, and Miguel, 2012; Dunning et al., 2019b; Ioannidis, 2005). Second, government programs' impact may differ from academic or non-governmental ones (Vivalt, 2020). Third, external validity, both in transporting and scaling up policy, poses significant challenges (Williams, 2020). Finally, impact evaluations may lack the granularity relevant to front-line service providers.

the cultural habit of indulging in conversations about different types of mangoes and their use during that period of the year.

Both treatment and control groups subsequently underwent refresher training encompassing antenatal care, HIV prevention and testing, and pregnancy-related danger signs. These topics were strategically chosen to enhance CHWs' impact on lagging health indicators at the local level, specifically the share of pregnant women completing at least four antenatal checks and the share of women giving birth assisted by skilled health personnel. We designed visual aids and a scripted module aligned with existing program training materials to facilitate the training. Drawing on their roles as former program supervisors, trainers emulated the established style and structure of routine monthly refresher training sessions.

3.2 Randomization and Sampling

The intervention was implemented in partnership with the Ministry of Public Health of Guinea-Bissau in the rural administrative region of Oio, constituted by the health regions of Farim e Oio.[23] The Farim health region includes five health areas, while the Oio health region has ten. Each health area typically represents the territory covered by a health center. In January 2022, the Farim health region contained 93 villages with at least one CHW and a total of 139 CHWs; the Oio health region had 286 villages with at least one CHW and a total of 463 CHWs. This results in a combined 379 villages with at least one CHW, totaling 602 CHWs. All villages with at least one CHW were eligible. I sampled one CHW per village, resulting in 379 CHWs being invited to participate in the study. Eventually, 247 CHWs participated in workshops, constituting 41 percent of CHWs active in the two health regions.

I assigned selected villages to either *Treatment* and *Control* - figure D1 in Appendix D summarizes the assignment to treatment. I created pairs of villages stratifying on health area, distance from the health center[24], and number of CHWs in the village.[25] I then randomly split each pair of villages between *Treatment* and *Control*, resulting in 180 villages assigned to the treatment group and 190 villages to the control group. Within the village, I randomly ranked CHWs. The first-ranked CHW within the village was selected to participate in the study. The remaining ones were labeled as

[23]Guinea-Bissau is divided into eleven health regions and nine administrative regions.

[24]The specific variable distinguish distances of less than five km, between five and ten km, between ten and twenty km, and more than twenty km.

[25]The specific variable distinguishes villages with only one and villages with more than one CHW.

substitutes according to their ranking. Sixty-five percent of villages in the sample had only one CHW and no replacement. Research assistants based in Bissau, the country's capital city, reached out to CHWs following a pre-defined calling protocol. The script to invite CHWs to participate in the study was identical for CHWs assigned to treatment and control groups. 133 CHWs from treatment villages and 114 CHWs from control villages participated in the workshop.

When invited to participate in the study, CHWs were also requested to compile a list of 20 households among their assigned ones. The only recommendation was to select them based on the extent to which CHWs anticipated the household could benefit from maternal and child health promotion. Of the 247 CHWs who participated in the workshop, 233 delivered the household list. Hence, all CHWs produced these lists before exposure to treatment. I then randomly selected a sample of households from these lists - see graph D2.

3.3 Measurement

The measurement in this project includes a broad range of data sources. These encompass two phone surveys and several face-to-face data collection activities.[26] For each phone survey, I set up one phone survey team with the ability to speak in local languages based in Bissau, the country's capital city. The phone surveys include (i) a households' endline phone survey conducted six months after the intervention and (ii) a CHWs' endline phone survey conducted four months after the intervention. For the face-to-face data collection activities, I set up two survey teams of three enumerators each who, together with the trainers, implemented the workshops. The face-to-face data sources include (iii) CHWs' pre and post-intervention face-to-face beliefs elicitation, (iv) CHWs' pre and post-intervention face-to-face surveys, and (v) CHWs' pre and post-intervention pen-and-paper knowledge tests. I now turn to provide some details about the design of these data.

The household phone survey I designed and conducted was administered between December 2, 2022, and January 17, 2023. I randomly assigned households to enumerators stratifying based on the sampling strata, including the treatment status of the respective CHW, which was concealed from the enumerators. Only women between 18 and 49 years old were eligible to be interviewed, as they

[26] Recent empirical findings highlight differences in the statistical characteristics of phone responses compared to in-person measures. While phone responses display higher mean and variance, treatment effect estimations remain stable across survey mode. Notably, phone data estimates are less precise than other modes (Anderson et al., 2023).

represent the CHWs program's target population.[27] A total of 827 respondents were interviewed. The questionnaire's main focus is checking whether the respondent knows the CHW assigned to her household and assessing the quantity and quality of healthcare services delivered by CHWs. The relevant periods assessed are the past month, explicitly referred to as November 2022, and the past six months, explicitly fixing the start of the period in June 2022.[28] Additionally, it elicits the distance in minutes from the CHW's house.[29] Moreover, the questionnaire includes demographic questions about the respondent, such as education, religion, and civil status, and about her household's size and composition, focusing on the presence of pregnant women or children under 5. Finally, it contains a test to measure downstream outcomes on knowledge of health during pregnancy and antenatal care.

The CHWs phone survey I designed and conducted was administered between October 19 and November 3, 2022, four months after the intervention. I randomly assigned CHWs to enumerators stratifying based on the treatment group, which was concealed from the enumerators. A total of 236 CHWs were interviewed. The questionnaire includes questions to assess their intentions regarding the quantity and quality of healthcare services provision in the next month, explicitly referred to as November 2022. I designed the data collection timeline to leave November 2022 as the reference time period in the CHW and household phone surveys. This way, I can directly compare CHWs' self-reported service delivery intentions with the households' assessment of the healthcare services received over the same period. The questionnaire, moreover, contains a test to assess CHWs' knowledge of health during pregnancy and antenatal care.

The CHWs' pre and post-intervention face-to-face beliefs elicitation I designed and conducted took place during the workshops between May 17 and June 5, 2022. CHWs were randomly assigned to one of the three enumerators comprising the team upon arrival, which they kept for all the face-to-face survey modules during the workshop. I elicit two sets of participants' beliefs before and after the intervention. The first set of beliefs I elicit are prior and posteriors about the status of six of the program's target health indicators at the local level. These six indicators comprise the three

[27]The CHWs program focuses on children under 5, pregnant women, recent mothers, and, more broadly, women within the reproductive age range of 15 to 49. I consider this last category for eligibility and impose being of legal age as a requirement for informed consent in survey participation.

[28]To mitigate recalling bias, enumerators further characterized June as the beginning of the rainy season, which in Guinea Bissau generally goes from June to October.

[29]To keep measurement homogeneous, enumerators ask respondents to estimate the distance from a precise location, which they do not label as the CHW's house. They instead identify it using information such as street names and landmarks previously provided by the CHWs to enable someone to locate their house.

indicators object of the treatment module, namely i) under-five mortality rate, ii) share of pregnant women completing at least four antenatal checks, and iii) share of women giving birth assisted by skilled health personnel. Moreover, I add three more indicators, namely, iv) the share of women giving birth at a health facility, v) the share of women completing all recommended post-natal checks, and vi) the share of newborns receiving all vaccines on time. I design and employ a modified ruler with a scale from 0 to 100 as a slider to elicit subjective beliefs precisely - see Figure E1 in the Appendix.[30]. The second set of beliefs I elicit is, for each health indicator, the CHWs' subjective perception of the program's effectiveness in improving each referred health indicator. I measure answers on a Likert scale from 1 to 5 with anchors 1 (Not at all) and 5 (Extremely).[31]

The pre-intervention face-to-face survey includes questions on the CHWs' demographic and socioeconomic characteristics. Moreover, it assesses the quantity of healthcare services delivered by CHWs in the past month, explicitly referred to as April 2022. The post-intervention face-to-face survey includes additional questions on the CHWs' demographic and socioeconomic characteristics. Moreover, it elicits CHWs' intentions regarding the quantity of healthcare services provision in the next month, explicitly referred to as June 2022. Finally, it assesses several dimensions of motivation through a psychometric module adapted from validated survey modules. In particular, I measure mission motivation (Khan, 2020), identity and self-image concerns (Hackman and Oldham, 1975), and warm-glow motivation (Grant, 2008).

The pen-and-paper self-administered pre and post-training tests I design and conduct assess participants' knowledge of prenatal care. The pre-training test was administered at the beginning of the workshop, before the intervention and the refresher training. The post-training test was administered at the end of the workshop, after both the intervention and the refresher training. They are based on the CHWs program material and include 26 questions, mixing multiple-choice and true or false questions. While all questions focus on health during pregnancy and antenatal care, the tests focus on the three main topics covered in the refresher training: antenatal checks, HIV prevention and testing, and danger signs during pregnancy.[32]

[30]See Delavande, Giné, and McKenzie (2011) for a discussion on using visual aids to elicit probabilistic expectations in developing countries

[31]Appendix Section E.1 reports details about the beliefs elicitation exercise.

[32]Appendix Section E.2 presents the full pen-and-paper knowledge test. Appendix Section E.3 reports the over-the-phone test included in the household and CHWs phone surveys, a reduced and adapted version of the pen-and-paper one. To summarize the full knowledge test and optimize it for communication over the phone, I run a PCA for the control group's post-training pen-and-paper test grades, select the first 10 questions of the

4 Estimation strategy

The experiment is designed to study the impact of providing CHWs with correct and up-to-date evidence of their program's effectiveness in achieving its mission. I am mainly interested in assessing the impacts on CHWs' performance, as reported in the Pre-analysis Plan (PAP), registered for this project at the American Economic Association (AEARCTR-0009054), titled 'Information Design To Improve Service Delivery: Experimental Evidence from Guinea-Bissau.' Hence, the main hypotheses are the following.

Hypothesis 1. The intervention improves service delivery. Households assigned to CHWs in the treatment group receive more and better quality healthcare services than those assigned to the control group.

Hypothesis 2. The intervention leads CHWs in the treatment group to exert more effort and improve performance compared to the control group.

Hypothesis 2 a. CHWs in the treatment group learn more from the refresher training and show larger improvements in grades than the control group after the refresher training.

Hypothesis 3. The intervention leads CHWs in the treatment group to change their perceptions of the program's effectiveness in fulfilling its mission relative to those in the control group.

Hypothesis 3 a. CHWs in the treatment group achieve more accurate posterior beliefs on the status of the program's target indicators than the control group.

Hypothesis 3 b. CHWs in the treatment group update their beliefs on the program's contribution to improving the target indicators.

To evaluate the main hypotheses, I estimate a set of specifications. The first specification I consider is the following.

$$y_i = \alpha + \beta_1 \text{Treat}_i + Z_i'\gamma + \epsilon_i \tag{1}$$

Where y_i is the outcome of interest at the follow-up, i.e., related to CHW or household outcomes. Note that individual i can be a CHW or a woman (18-49 years old) belonging to a household assigned

first PCA component, and transform the selected single-choice questions into 'True or False' statements.

to that CHW. Treat_i, the treatment indicator, is a binary variable taking value 1 for CHWs or households whose CHWs were assigned the corresponding treatment. Z_i is a set of controls, including strata fixed effects, defined by the blocks formed before randomization.[33] ϵ_i is an idiosyncratic error term assumed to be clustered at the CHW level.

When pre-intervention values of the outcome variable are available, I employ a Difference-in-Difference model, given the high auto-correlation of the outcomes of interest. The specification I consider is the following.

$$y_i = \alpha + \beta_1 \text{Treat}_i + \beta_2 \text{Post}_i + \beta_3 \text{Treat}_i \times \text{Post}_i + Z_i'\gamma + \epsilon_{ib} \tag{2}$$

Where y_i is the outcome of interest at the follow-up, i.e., related to CHW outcomes, Treat_i is a binary variable taking a value of 1 if the CHW is assigned to the treatment group, Post_i is a binary variable taking a value of 1 for post-intervention observations, Z_i' is a set of controls, including strata fixed effects, and when appropriate, individual fixed effects.[34] The coefficient of interest is β_3. ϵ_i is an idiosyncratic error term assumed to be clustered at the CHW level.

I investigate heterogeneous effects by estimating the following specification.

$$y_i = \alpha + \beta_1 \text{Treat}_i \times \text{X}_i + \beta_2 \text{Treat}_i \times (1 - \text{X}_i) + X_i\delta + Z_i'\gamma + \epsilon_{ib} \tag{3}$$

Where X_i is an indicator of where some relevant worker's baseline characteristic or pre-intervention values of the outcome variable stand compared to the sample median. The coefficients of interest are β_1 and β_2, which capture, respectively, the effect of Treat_i on workers with $\text{X}_i = 1$ (above the median) and $\text{X}_i = 0$ (below or equal to the median). Z_i' is a set of controls, including strata fixed effects.[35] ϵ_i is an idiosyncratic error term assumed to be clustered at the CHW level.

For all specifications, I estimate linear regressions regardless of whether the outcomes are con-

[33] Control variables include CHWs' characteristics (education, religion, employment in the past twelve months) and women's characteristics (education and religion). Strata fixed effects include those used for CHWs' treatment allocation and households' sampling, as presented in Section 3.2.

[34] Control variables include CHWs' characteristics (education, religion, employment in the past twelve months). Strata fixed effects include those used for CHWs' treatment allocation, as presented in Section 3.2.

[35] Control variables include CHW characteristics (education, religion, employment in the past twelve months). Strata fixed effects include those used for CHWs' treatment allocation, as presented in Section 3.2.

tinuous or discrete.[36] The outcome variables and the dimensions for heterogeneous analysis considered in this paper are included in the PAP. To adjust the inference for multiple hypotheses testing and control for overall false rejections among reported outcomes in each table, I also provide "sharpened" False Discovery Rate (FDR) – adjusted q-values following the procedure of Benjamini, Krieger, and Yekutieli (2006), as outlined by Anderson (2008).

5 Results

5.1 Balance and descriptive statistics

I show balance tests in Section F of the Appendix. For CHWs, I check the balance on CHWs' baseline demographic characteristics and background as CHWs (Table F1). I find only one unbalanced characteristic. Namely, the CHWs assigned to the treatment group are, on average, less likely to have had a paid job in the 12 months before the survey. I also check the balance of baseline beliefs on health indicators (Table F2) and baseline perceptions of the CHW program's contribution to the evolution of each health indicator (Table F3). I do not find any unbalance for these variables. I can also employ Table F1 to describe CHWs' characteristics by looking at the control group's mean. The CHWs' average age is 43. Four percent are female, 75 percent are household heads, and 47 percent are Muslim. Seventy-six percent of participants completed more than six years of education, and 11 percent are currently studying. Seventy-nine percent had a paid job during the past 12 months, and 35 percent owned a business at the time of the survey. Ninety-seven percent of participants have already received training as CHWs. On average, each CHW covers almost two villages and has 60 households assigned. Participants report devoting 12.5 hours per week to their CHW activities, having visited 65 percent of their pre-assigned households in the past month, i.e., April 2022.

For program recipients, I check the balance for households and respondents' characteristics (Table F4). I run ten tests and find statistical significance for two variables. Namely, the respondents from households assigned to the treatment group are, on average, more likely to be Muslim and more

[36]The specifications I describe were registered in the PAP, where I also included an ANCOVA specification. In the analysis, I give preference to the pre-registered difference-in-differences specification given the high autocorrelation of outcomes variables in my sample, as reported at the bottom of the relevant results tables. In fact, ANCOVA maximizes statistical power in field experiments when the autocorrelation of the outcome variables is low (McKenzie, 2012).

likely to have completed six or more years of education. Appendix Table F4 also provides an overall characterization of the demographic characteristics of the households interviewed over the phone. All respondents are women, and the average age is 29 years.[37] Eleven percent are the household head, 77 percent are married, and 21 percent completed six or more years of education. Forty-three percent are Muslim. Households are composed of 14 members on average. Fifty-three percent of households have a pregnant woman, and 95 percent have a child under 5. The average distance from the assigned CHW's house is twelve minutes.

5.2 Treatment effects

5.2.1 Service delivery

This paper studies the causal impact on performance of providing workers with correct and up-to-date evidence of their program's effectiveness in fulfilling its mission. The main outcome of interest is healthcare service delivery through door-to-door household visits, the key indicator of CHWs' performance. Table 1 shows estimates of treatment effects on healthcare service delivery, measured as the number and duration of door-to-door visits. It employs Equation 1 and phone survey data at the level of the household. I find an overall positive treatment effect consistent with Hypothesis 1. The households accompanied by CHWs assigned to the treatment group report receiving 11 percent more visits during the past month, i.e., November 2022, significant at the 5 percent level (column 1). These visits lasted, on average, 22 percent longer, significant at the 5 percent level (column 2). Overall, these households received 39 percent more time in healthcare services delivered by their CHW compared with households accompanied by a CHW in the control group, significant at the 1 percent level (column 3). Considering the six months between the intervention and the survey, i.e., from June to November 2022, the results are comparable. Households accompanied by a CHW in the treatment group received 18 percent more healthcare time than households accompanied by a CHW in the control group, significant at the 5 percent level (column 6). This effect comes mostly from the 14 percent longer duration of visits, significant at the 5 percent level (column 5), as I do not observe any significant treatment effect on the number of visits (column 4). With the increased healthcare received, households accompanied by a CHW in the treatment group show a 3 percent increase in

[37] Note that only women between 18 and 49 years old, the program's target recipients, were eligible to be interviewed.

satisfaction with the program, significant at the 10 percent level (column 7).[38] Table D1 presents ex-ante and ex-post minimum detectable effect size estimates for the outcomes presented in Table 1.

The increased healthcare time delivered to households may be compensated by CHWs targeting households living closer to their house or not fitting the program's target population. Employing Equation 1 and phone survey data at the level of the household, I find no evidence of a treatment effect on targeting by distance or the presence in the household of a pregnant woman or a child under 5 in the last month (Table H1) or the past six months (Table H2).

During the additional time spent delivering healthcare service to households, CHWs in the treatment group promoted more health practices, the core task of door-to-door household visits, consistent with Hypothesis 1. Table 2 reports which health practices households recall being promoted by the CHW during the last visit received, employing Equation 1 and phone survey data at the level of the household. The practices are aggregated in indexes referring to specific health dimensions. Compared to the control group, households assigned to CHWs in the treatment group are 3 percent more likely to recall receiving health promotion referring to child care, significant at the 5 percent level (column 2), and preventive or treatment practices for diarrhea, malaria, and pneumonia, significant at the 10 percent level (column 3). I do not see any significant change in the promotion of practices referring to 'Perinatal care' (column 1), 'HIV, TBC' (column 4), and 'Preventive practices at home' (column 5). It is important to note that the control group's mean is already high, leaving little room for improvement.

I leverage the data collection design and use November 2022 as the reference time period to make a direct comparison between the healthcare services households report having received and CHWs' self-reported intentions to deliver healthcare services over the same period. Employing Equation 1 and phone survey data at the level of the CHW, I find an overall positive treatment effect consistent with Hypothesis 2. Table 3 shows that CHWs assigned to the treatment group report the intention to spend 27 percent more time with households during the next month, significant at the 10 percent level, computed by the number of visits planned times the expected average duration of a single visit (column

[38]Table J1 in the Appendix presents estimates of the heterogeneous treatment effects on service delivery by the household sampling strata (see graph D2). Importantly, I do not observe significant heterogeneity when considering the households flagged in the nudging exercise included in the CHWs' phone survey. In this sense, the effect of the nudge does not appear to be significantly different across treatment groups. The only heterogeneous effect I observe is for the stratum identifying those households contacted by CHWs during the health promotion activity performed at the workshop. These households received, on average, 36 percent more visits over the past month.

3). This effect comes from a 24 percent increase in the average expected duration of a single visit, significant at the 10 percent level (column 2). It is worth noting the consistency between the expected average duration of a single visit stated by CHWs and the realized one reported by households. Just focusing on the control group, CHWs report the intention of carrying out visits of an average duration of 15.23 minutes in November 2022. Employing phone survey data at the level of the household, those assigned to a CHW in the control group report an average duration of the last visit received of 15.07 minutes. It is reassuring to see the CHWs' self-reported intentions confirmed by program recipients.

5.2.2 Learning about health practices

Let us turn to the treatment effects on learning about health practices. Both treatment and control groups, after receiving either the intervention or the placebo, underwent refresher training encompassing antenatal care, HIV prevention and testing, and pregnancy-related danger signs. Table 4 shows estimates of treatment effects on CHWs' learning of the health practices covered in the refresher training, employing Equation 2 to compare grades in the pre and post-training knowledge tests. I find an overall positive treatment effect consistent with Hypothesis 2a. CHWs in the treatment group show an overall 6 percent larger grade improvement between pre and post-training knowledge tests, significant at the 10 percent level, compared to the control group (column 1). The different test components show that this effect is concentrated in the module assessing CHWs' knowledge of 'Danger signs in pregnant women' (column 4). The magnitude of the effect is 14 percent, significant at the 5 percent level. I do not find significant treatment effects on the other components of the test, namely 'Antenatal care' (column 2) and 'HIV prevention and testing' (column 3).[39]

5.2.3 Beliefs about the organization's effectiveness

Let us now turn to how workers' beliefs respond to evidence of the organization's effectiveness in fulfilling its mission. Consistent with Hypothesis 3, I find that the intervention leads workers in the treatment group to update their beliefs.

[39]I present additional results on learning about health practices in the Appendix. By employing the reduced version of the knowledge test, Table I1 shows the treatment effects on learning at the level of the CHW up to four months after the intervention, while Table I2 presents estimates of treatment effects on knowledge diffusion at the level of the household. I also report balance tables for the pre-training knowledge test grades, both for the full test (Table I3) and the reduced version (Table I4). I note an overall unbalance, with the CHWs in the treatment group obtaining lower grades in the pre-training knowledge test.

First, let us devote our attention to how workers update their beliefs about the program's target health indicators at the local level. Table 5 shows estimates of treatment effects, in standard deviations, on the change in CHWs' beliefs inaccuracy, computed as the absolute value of the difference between the elicited belief and the health indicator's value. It employs Equation 2 to compare pre and post-intervention beliefs data at the level of the CHW.[40] I find an overall positive treatment effect on belief accuracy. The posterior beliefs of treated CHWs are consistently closer to the correct value across all the health indicators included in the beliefs elicitation exercise, consistent with Hypothesis 3a. The treatment group, compared to the control group, shows a 20 percent decrease, corresponding to 0.33 σ, in the average error across indicators, significant at the 1 percent level (column 1). Indicators presented in columns 2, 3, and 4 were directly addressed in the intervention, while indicators in columns 5, 6, and 7 were not.[41] Looking at each indicator individually, I see that relative to the control group, the treatment group's average error decreased by 18 percent, or 0.37 σ, for the share of births attended by qualified healthcare personnel (doctors, nurses, or trained midwives), significant at the 1 percent level (column 4); 21 percent, or 0.48 σ, for the share of births that took place at the health center, significant at the 1 percent level (column 5); 23 percent, or 0.41 σ, for the share of new mothers who completed all recommended postnatal checks, significant at the 1 percent level (column 6); and 14 percent, or 0.31 σ, for the share of newborns who all the recommended vaccines on time, significant at the 5 percent level (column 7). I do not see a significant treatment effect on belief accuracy for the under-5 mortality (column 2) and the percentage of pregnant women who completed the four recommended prenatal checks before delivery (column 3). However, the magnitude and direction of the effects are consistent with the other indicators.[42]

Second, let us turn to the direction of belief updating. Figure 1 shows estimates of treatment effects, in standard deviations, on the change in the CHWs' beliefs before and after the intervention. It employs Equation 3 to investigate the relevance of priors for how CHWs update their beliefs on the status of health indicators at the local level in response to the intervention, employing pre and post-intervention beliefs data at the level of the CHW.[43] I find that the treatment overall leads to downward

[40]The choice of a difference-in-differences approach is motivated by the high autocorrelation of pre and post-intervention beliefs, as presented in the last row of Table 5.

[41]Figure **??** in the Appendix shows the correlation of prior beliefs about health indicators.

[42]Table G1 in the Appendix presents estimates of treatment effects on the change in the difference between CHWs' beliefs and the health indicator's value.

[43]Figure G1 in the Appendix shows that the treatment is particularly effective at correcting the more inaccurate beliefs.

updating for most indicators from pre to post-intervention beliefs. The treatment group, compared to the control group, shows a 0.43 σ reduction in the average value of their beliefs about the status of health indicators at the local level, significant at the 1 percent level (Panel A). I observe a different treatment effect on belief updating for the health indicators directly addressed in the intervention compared to those that were not. For indicators included in the intervention, the treatment leads the CHWs with relatively optimistic priors to significantly update their beliefs compared to the control group. In fact, CHWs holding prior beliefs above the median reduce them by 0.57 σ more than the control group, significant at the 1 percent level, while those holding prior beliefs below or equal to the median do not show any significant updating (Panel B). For indicators not included in the intervention, on the other hand, prior beliefs do not appear to be as relevant. In fact, both CHWs with relatively optimistic and pessimistic priors respond to the intervention by updating their beliefs downward relative to the control group. In fact, I observe that CHWs holding prior beliefs above the median reduce them by 0.66 σ, significant at the 1 percent level, and those holding priors below or equal to the median reduce them by 0.48 σ, significant at the 5 percent level (Panel C).

Finally, let us look at how workers update their perceptions of the program's effectiveness in improving health indicators at the local level. Figure 2 shows estimates of treatment effects, in standard deviations, on the change in the CHWs' perceived program's effectiveness before and after the intervention.[44] It employs Equation 3 to investigate the relevance of priors for how CHWs change their perceptions about the program's effectiveness in response to the intervention, employing pre and post-intervention beliefs data at the level of the CHW.[45] I find that the treatment overall leads to a change in the perceived program's effectiveness in improving health indicators at the local level, consistent with Hypothesis 3b. The direction of this change is negative, with an overall reduction in the perceived effectiveness of the program. The treatment group, compared to the control group, shows a 0.18 σ reduction in the average value of their beliefs about the program's effectiveness in improving health indicators, significant at the 5 percent level (Panel A). I observe a different treatment effect on the change in perceptions about the program's effectiveness in improving the health indicators that were directly addressed in the intervention compared to those that were not. For indicators included in the intervention, priors do not seem to matter. I observe no significant treatment effect on belief

[44]Table G2 in the Appendix presents estimates of treatment effects on the change in the CHWs' perceived program's effectiveness after the intervention, while Table G3 adds the follow-up four months later.

[45]Figure **??** shows the correlation of prior beliefs about the program's effectiveness.

updating compared to the control group (Panel B). For indicators not included in the intervention, on the other hand, the treatment leads both CHWs with relatively optimistic and pessimistic priors about the program's effectiveness update downward relative to the control group. In fact, I observe that CHWs holding prior beliefs above the median reduce them by 0.74 σ, significant at the 1 percent level, and those holding priors below or equal to the median reduce them by 0.42 σ, significant at the 5 percent level (Panel C).[46]

5.2.4 Motivation

The intervention, besides changing workers' perceptions of the organization's effectiveness in fulfilling its mission, may be altering their motivation. Table 6 presents estimates of treatment effects, in standard deviations, on potentially relevant motivation dimensions. It employs Equation 1 and face-to-face survey data at the level of the CHW. Overall, I find no significant treatment effect on any of the dimensions considered.[47] First, I do not observe any significant treatment effect on mission motivation, measured as in Khan (2020). In particular, the CHWs in the treatment group are no more likely than those in the control group to enjoy the program because of the importance it places on the mission of improving health in your community, to feel more aligned with the program's mission, or to feel attached to the program because of it (column 1). Second, following Bénabou and Tirole (2006), I turn to the perceived compatibility of the position with CHWs' self-image. I do not observe a significant treatment effect on the perceived fitness of the position. In particular, CHWs in the treatment group are no more likely to perceive the program as a good fit for how they see themselves in the present or in the future than those in the control group (column 2). Closely connected, I look at the potential treatment effects on identity in the spirit of Akerlof and Kranton (2000, 2005). I do not observe a clear treatment effect on this dimension either. The CHWs in the treatment group are no more likely than those in the control group to like the identity given by the job or to feel that the activities related to the job satisfy their identity (column 3). Finally, I do not observe a significant treatment effect on perceived pleasure from helping program recipients, i.e., their warm-glow (Andreoni, 1990).

[46] Figure G2 in the Appendix presents the same estimates for the change in perceived effectiveness four months after the intervention.

[47] The lack of effects on the motivation indices may be partly due to their measurement. I assess motivation by measuring agreement with a set of statements through a 5-item Liket scale with anchors 1 'Strongly disagree' and 5 'Strongly agree'. I observe a relatively low variation in answers.

6 Cost-Effectiveness and scaling up

The intervention is cost-effective (Appendix Section K). I estimate that the intervention costs 2.91 USD per CHW. As a result, in my setting, increasing healthcare time delivered by CHWs by 1 standard deviation has a per capita cost of 5.56 USD. These results confirm the cost-effectiveness of leveraging existing evidence to motivate workers. It is worth noting some aspects of the intervention implementation that may be important to inform a possible scaling-up of the intervention. First, the intervention relies on existing evidence of the organization's effectiveness in achieving its mission. Hence, the content, scope, and granularity of the evidence available might influence the intervention design and implementation at scale. Second, the fact that the treatment effect is persistent over the six-month period following the intervention provides reasons for cautious optimism. The intervention may effectively motivate agents in the long run, although I cannot measure the effect past the experimental period. Third, CHWs were not aware of the existence of an experimental module when invited to participate in the workshop. This minimized concerns about selection effects and information spillovers to CHWs in the control group. However, at scale, we may expect a selection effect in the intervention: some CHWs may actively seek information, while others may prefer to avoid it. Finally, the study participants' activity as CHWs might not be their main occupation. Therefore, the intervention's effects might show seasonal fluctuations depending on the opportunity cost of time and effort allocated to the CHW program activities.

7 Conclusion

This paper provides the first experimental evidence that an organization can leverage existing evidence of mission accomplishment to enhance performance without altering incentives. I report on the results of a field experiment testing an evidence-provision intervention in collaboration with a nationwide public health program of the Ministry of Public Health of Guinea-Bissau. In particular, I provide front-line service providers with correct and up-to-date evidence of the program's effectiveness in achieving the claimed mission of improving maternal, newborn, and child health. The intervention draws upon the official program evaluation and provides data on the program's target health indicators at the local level, their recent evolution, and the program's effectiveness in improving them.

I show the feasibility of leveraging existing evidence of mission accomplishments to enhance

workers' performance even in settings with scarce resources. The main finding is the improvement in service delivery six months after the intervention. In particular, program recipients assigned to workers in the treatment group report receiving more frequent and longer door-to-door visits compared to those in the control group. Moreover, they note a more comprehensive healthcare promotion and express higher satisfaction with the program. The increase in performance can be interpreted through the effects of the intervention on beliefs and effort. Firstly, the intervention effectively corrected prior beliefs among workers in the treatment group regarding the status of local health indicators, leading to more accurate posterior beliefs than the control group. This adjustment predominantly resulted from workers with initially optimistic priors revising their beliefs downward. Secondly, the intervention prompted workers to adjust their perceptions of the program's effectiveness in improving health indicators, reducing the treatment group's average perception of mission accomplishment. Importantly, this shift in beliefs is followed by an observed increase in effort compared to the control group. Workers in the treatment group appear to learn more during skill-refresher training and achieve higher grades than the control group in a subsequent test assessing knowledge of health practices.

These findings hold significant implications for public service delivery programs and underscore the potential of evidence-driven strategies to advance organizational goals. Sharing evidence with employees emerges as an effective strategy to enhance performance and improve service delivery. It is worth noting that these effects may be sensitive to the content and framing of the evidence shared and its interaction with existing priors. In any case, the direct policy recommendation emanating from this study is for mission-oriented organizations to explore the untapped potential of available evidence.

Main Tables

Table 1: Service delivery at the household level

	HHs phone survey						
	Past month			Past six months			
	Visits received	Visit duration (minutes)	Time with the CHW (minutes)	Visits received	Visit duration (minutes)	Time with the CHW (minutes)	Satisfaction with the program
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Treat	0.145**	3.321**	10.253***	0.014	3.332**	24.188**	0.112*
	(.074)	(1.313)	(3.131)	(.280)	(1.332)	(11.900)	(0.061)
	[.042]	[.027]	[.008]	[.159]	[.027]	[.042]	[.047]
Control Mean	1.32	15.07	26.52	5.27	23.52	132.71	4.158
Observations	786	782	782	789	786	786	795
R^2	0.156	0.129	0.125	0.137	0.156	0.205	0.075
Effect (%)	11.01	22.04	38.66	0.27	14.17	18.23	2.69

Note: In this table, I report the intent-to-treat estimates of the direct effects of the treatment on service delivery to households. These are obtained by ordinary least squares (OLS) estimation of Equation 1. The estimation sample consists of women (18-49 years old) interviewed in the follow-up phone survey. Depending on the column, the dependent variables are defined by the following. (1) Number of visits received during November 2022. (2) Average duration of a visit during November 2022. (3) The number of visits during November 2022 multiplied by the average duration of a visit during November 2022. (4) Visits received since June 2022. (5) Average duration of a visit since June 2022. (6) The number of visits since June 2022 multiplied by the average duration of a visit since June 2022. (7) A 5-item Likert scale that takes the values 1 'Not at all,' 2 'A little,' 3 'Moderately,' 4 'A lot,' and 5 'Extremely' to measure replies to the following question: 'How satisfied are you with the CHWs program?' All specifications include CHW strata fixed effects. They also include CHW and household-level controls. The full list of controls is presented in Section 4. Standard errors are reported in parentheses and q-values in brackets, obtained using the sharpened procedure of Benjamini, Krieger, and Yekutieli (2006) at the table level. Standard errors are clustered at the CHW level. *** p<0.01, ** p<0.05, * p<0.1.

Table 2: Health practices promoted by CHW during the last visit

	HHs phone survey				
	Perinatal care	Child care	Diarrhea, Malaria, Pneumonia	HIV, TBC	Preventive practices at home
	(1)	(2)	(3)	(4)	(5)
Treat	0.001	0.032**	0.025*	0.005	-0.008
	(.015)	(.013)	(.014)	(.028)	(.012)
	[1]	[.163]	[.244]	[1]	[1]
Control Mean	.93	.92	.92	.80	.96
Observations	797	797	797	796	797
R^2	0.053	0.066	0.071	0.071	0.061
Effect (%)	.11	3.48	2.71	.62	-.83

Note: In this table, I report the intent-to-treat estimates of the direct effects of the treatment on health practices promotion during household visits. These are obtained by ordinary least squares (OLS) estimation of Equation 1. The estimation sample consists of women (18-49 years old) interviewed in the follow-up phone survey. Depending on the column, the dependent variables are defined by the following. (1) Perinatal care index: index variable averaging two indicator variables for the following health practices: i) prenatal checks for the pregnant woman and ii) post-natal care for the newborn and the mother. (2) Child care index: index variable averaging three indicators variables for the following health practices: i) breastfeeding, ii) child malnutrition, and iii) child vaccination. (3) Treatment practices index: index variable averaging three indicators variables for the following health practices: i) diarrhea prevention and treatment, ii) malaria prevention and treatment, iii) pneumonia prevention and treatment. (4) Monitoring practices index: index variable averaging three indicators variables for the following health practices: i) HIV prevention, ii) tuberculosis prevention, and searching. (5) Preventive practices index: index variable averaging three indicators variables for the following health practices: i) impregnated mosquito nets use, ii) hand washing, iii) water treatment with bleach. All specifications include CHW and household-level strata fixed effects. They also include CHW and household-level controls. The full list of controls is presented in Section 4. Standard errors are reported in parentheses and q-values in brackets, obtained using the sharpened procedure of Benjamini, Krieger, and Yekutieli (2006) at the table level. Standard errors are clustered at the CHW level. *** p<0.01, ** p<0.05, * p<0.1.

Table 3: Self-reported service delivery over next month

	CHWs phone survey		
	Visits planned (of assigned)	Visit duration (minutes)	Time with households (minutes)
	(1)	(2)	(3)
Treat	0.020	3.606*	211.693*
	(0.033)	(2.167)	(119.779)
	[.222]	[.171]	[.171]
Control Mean	0.852	15.234	796.109
Observations	232	232	231
Effect (%)	2.35	23.67	26.59

Note: In this table, I report the intent-to-treat estimates of the direct effects of the treatment on self-reported intentions to deliver healthcare services to households. These are obtained by ordinary least squares (OLS) estimation of Equation 1. The estimation sample consists of CHWs interviewed in the follow-up phone survey four months after the intervention. Depending on the column, the dependent variables are defined by the following. (1) Number of household visits planned during November 2022 over the total number of households assigned. (2) Average expected duration of each household visit performed during November 2022. (3) Time planned to spend with households, computed by the number of visits planned multiplied by the average expected duration of each visit. All specifications include CHW-level strata fixed effects. They also include CHW-level controls. The full list of controls is presented in Section 4. Standard errors are reported in parentheses and q-values in brackets, obtained using the sharpened procedure of Benjamini, Krieger, and Yekutieli (2006) at the table level. Standard errors are clustered at the CHW level. *** p<0.01, ** p<0.05, * p<0.1.

Table 4: Change in grades from pre to post-training test - full test

	CHWs written tests			
	Full test	Antenatal care	HIV prevention and testing	Danger signs in pregnant women
	(1)	(2)	(3)	(4)
Treat X Post-test	0.041*	0.020	0.030	0.098**
	(0.021)	(0.023)	(0.033)	(0.042)
	[.094]	[.25]	[.25]	[.092]
Control Mean	0.679	0.666	0.680	0.706
Observations	489	489	489	489
R^2	0.423	0.339	0.322	0.350
Effect (%)	6.04	3.00	4.41	13.88
Control ρ	0.63***	0.56***	0.30**	0.44***

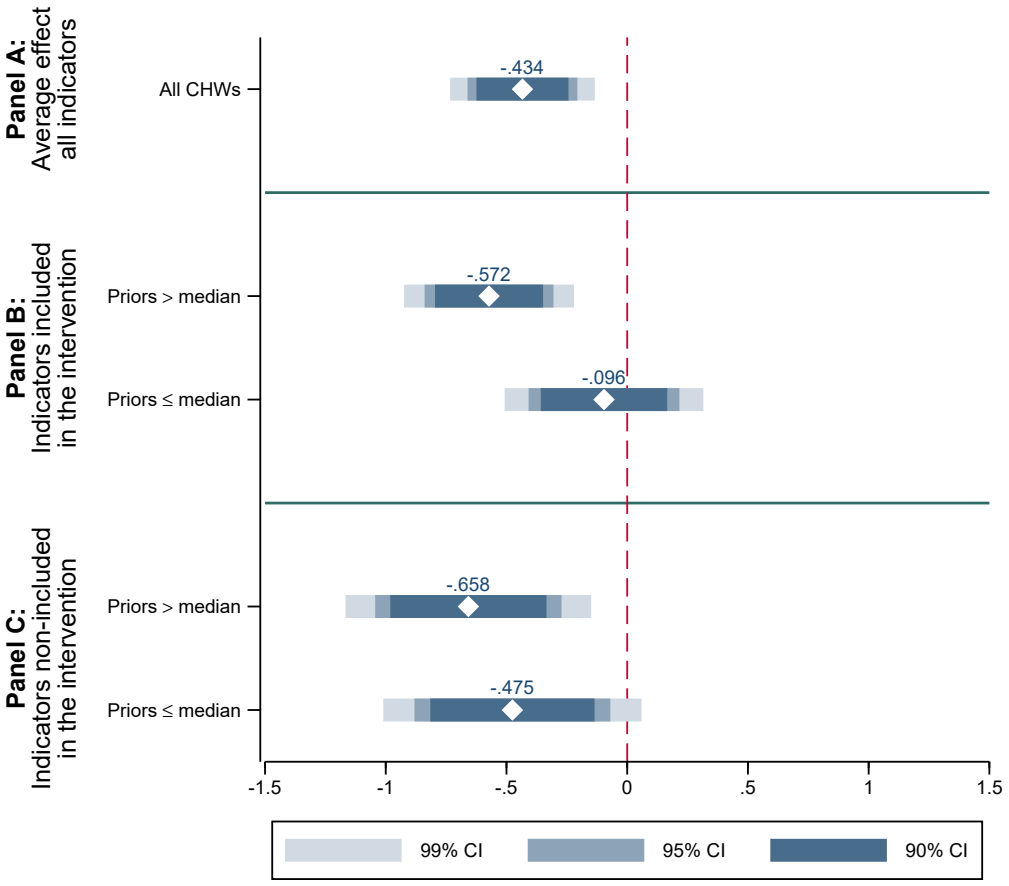
Note: In this table, I report the intent-to-treat estimates of the direct effects of the treatment on learning about health practices. These are obtained by ordinary least squares (OLS) estimation of Equation 2. The estimation sample consists of all the CHWs who took the post-intervention pen-and-paper test at the workshop. Depending on the column, the dependent variables are defined by the following. (1) Continuous variable indicates the share of correct answers over the entire test, containing 26 questions, and takes values from 0 to 1. (2) Continuous variable that indicates the share of correct answers to the 13 questions focused on antenatal checks during pregnancy, and takes values from 0 to 1. (3) Continuous variable that indicates the share of correct answers to the seven questions focused on HIV testing during pregnancy, and takes values from 0 to 1. (4) Continuous variable that indicates the share of correct answers to the six questions focused on the danger signs during pregnancy, and takes values from 0 to 1. All specifications include CHW-level strata fixed effects. They also include CHW-level controls. The full list of controls is presented in Section 4. Standard errors are reported in parentheses and q-values in brackets, obtained using the sharpened procedure of Benjamini, Krieger, and Yekutieli (2006) at the table level. Standard errors are clustered at the CHW level. *** p<0.01, ** p<0.05, * p<0.1.

Table 5: Distance between beliefs and health indicators

	CHWs face-to-face beliefs elicitation						
		Included in the intervention			Not included in the intervention		
	Average	U5 mortality	4 prenatal checks	Assisted birth	Birth at facility	Postnatal checks	Complete vaccines
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Treat X Post	-0.326***	-0.157	-0.228	-0.362***	-0.476***	-0.414***	-0.314**
	(.071)	(.118)	(.154)	(.133)	(.128)	(.126)	(.122)
	[.001]	[.057]	[.049]	[.008]	[.001]	[.002]	[.009]
Control Mean	.001	.000	.000	.000	.000	.000	.000
Observations	484	481	484	484	484	484	484
Effect (%)	-19.55	-24.56	-15.81	-17.87	-21.02	-23.33	-13.79
Control ρ	0.83***	0.73***	0.71***	0.67***	0.65***	0.64***	0.77***

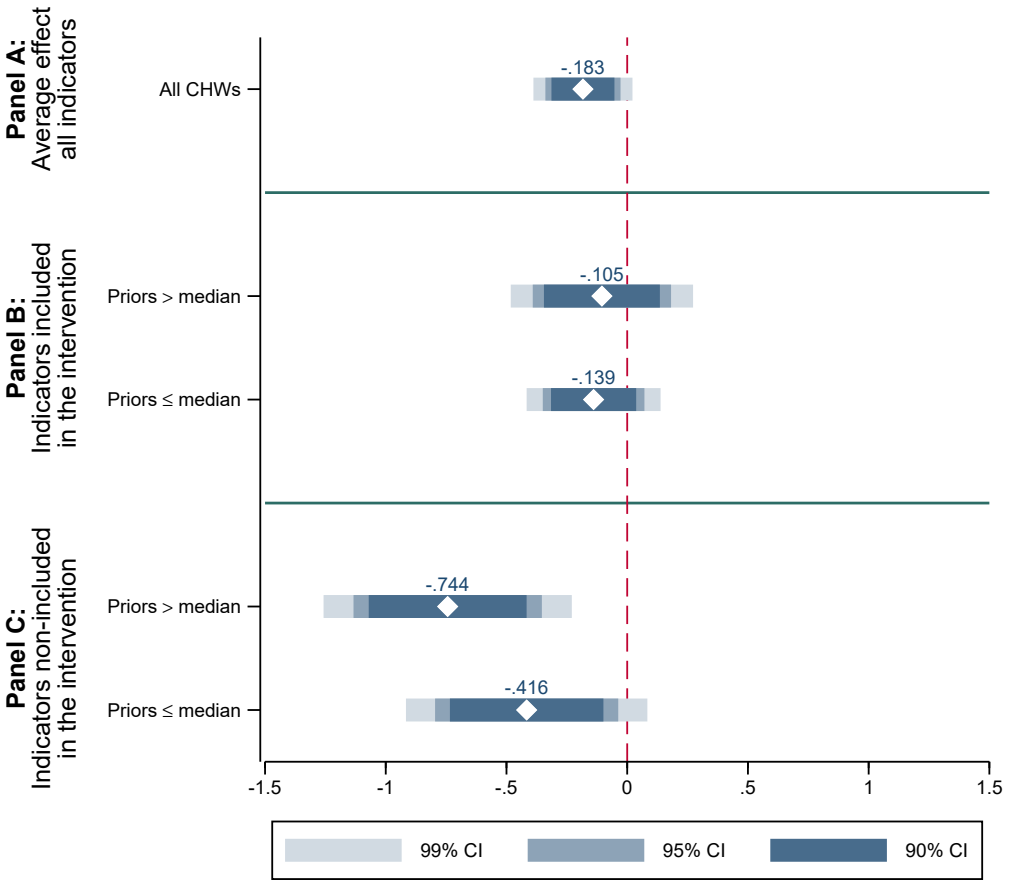
Note: In this table, I report the intent-to-treat estimates of the direct effects of the treatment on beliefs updating. These are obtained by ordinary least squares (OLS) estimation of Equation 2. The estimation sample consists of CHWs who participated in the pre and post-intervention face-to-face beliefs elicitation. In all columns, the dependent variable is a z-score representing belief inaccuracy, i.e., the absolute value of the difference between the elicited belief and the health indicator's correct value. The correct value is taken from MICS 2019. In column (1), the dependent variable is an index built using the procedure in Kling, Liebman, and Katz (2007). I calculate within-sample z-scores for each individual outcome, employing the mean and the standard deviation of the pure control group. I then obtain the unweighted average z-score for each group of outcomes. In columns (2)-(7), depending on the column, the relevant health indicator is described in the following. (2) 'In 2019, out of 100 children, how many would not survive to age 5?' (reversed). (3) 'In 2019, out of 100 pregnant women, how many completed the four recommended prenatal checks before delivery?'. (4) 'In 2019, out of 100 pregnant women, how many were giving birth attended by qualified health care personnel (doctors, nurses, or trained midwives)?'. (5) 'In 2019, out of 100 pregnant women, how many were giving birth at the health center?'. (6) 'In 2019, out of 100 new mothers, how many completed all recommended postnatal checks?'. (7) 'In 2019, out of 100 newborns, how many were receiving all the recommended vaccines on time?'. All specifications include CHW and household-level strata fixed effects. They also include CHW and household-level controls. The full list of controls is presented in Section 4. For each outcome, I report the mean outcome for the control group, the number of observations, the treatment effect in percentage points, and the autocorrelation (ρ) between pre and post-intervention outcome values. Standard errors are reported in parentheses and q-values in brackets, obtained using the sharpened procedure of Benjamini, Krieger, and Yekutieli (2006) a the table level. Standard errors are clustered at the CHW level. *** p<0.01, ** p<0.05, * p<0.1.

Figure 1: Effect of intervention on change in beliefs about health indicators - heterogeneous effects by priors



Note: Estimates based on OLS regressions using Equation 3. The estimation sample consists of CHWs who participated in the pre and post-intervention face-to-face beliefs elicitation. I define 'change in beliefs' as the difference between post and pre-intervention CHWs' beliefs about health indicators. In all panels, the dependent variable is an index built using the procedure in Kling, Liebman, and Katz (2007). I calculate within-sample z-scores for each individual outcome, employing the mean and the standard deviation of the pure control group. I then obtain the unweighted average z-score for each group of outcomes. Panel A plots the effect of the treatment on the average change in belief for all the six indicators elicited. For each indicator, pre and post-intervention beliefs are continuous variables that take values from 0 to 100. Panel B considers only the three indicators included in the intervention, described in the following: (1) 'In 2019, out of 100 children, how many would not survive to age 5?' (reversed); (2) 'In 2019, out of 100 pregnant women, how many completed the four recommended prenatal checks before delivery?'; and (3) 'In 2019, out of 100 pregnant women, how many were giving birth attended by qualified health care personnel (doctors, nurses, or trained midwives)?'. Panel C considers only the three health indicators not included in the intervention, described in the following: (1) 'In 2019, out of 100 pregnant women, how many were giving birth at the health center?'; (2) 'In 2019, out of 100 new mothers, how many completed all recommended postnatal checks?'; and (3) 'In 2019, out of 100 newborns, how many were receiving all the recommended vaccines on time?'. Both Panel B and C plot the treatment effect on the change in the average belief level for workers depending on their prior belief levels. The first bar in each panel reports the treatment effect for those CHWs holding priors above the median, while the second bar shows the treatment effect for those holding priors below or equal to the median. All specifications include CHW-level strata fixed effects. They also include CHW-level controls. The full list of controls is presented in Section 4. Standard errors are clustered at the CHW level.

Figure 2: Effect of intervention on change in perceived program's effectiveness - heterogeneous effects by priors



Note: Estimates based on OLS regressions using Equation 3. The estimation sample consists of CHWs who participated in the pre and post-intervention face-to-face beliefs elicitation. I define 'change in perceived program's effectiveness' as the difference between post and pre-intervention CHWs' beliefs about the effectiveness of the program in improving health indicators at the local level. In all panels, the dependent variable is an index built using the procedure in Kling, Liebman, and Katz (2007). I calculate within-sample z-scores for each individual outcome, employing the mean and the standard deviation of the pure control group. I then obtain the unweighted average z-score for each group of outcomes. Panel A plots the treatment effect on the change in the average perceived program's effectiveness in improving all the six indicators elicited. Panel B considers only the three indicators included in the intervention, described in the following: (1) 'In 2019, out of 100 children, how many would not survive to age 5?' (reversed); (2) 'In 2019, out of 100 pregnant women, how many completed the four recommended prenatal checks before delivery?'; and (3) 'In 2019, out of 100 pregnant women, how many were giving birth attended by qualified health care personnel (doctors, nurses, or trained midwives)?'. Panel C considers only the three health indicators not included in the intervention, described in the following: (1) 'In 2019, out of 100 pregnant women, how many were giving birth at the health center?'; (2) 'In 2019, out of 100 new mothers, how many completed all recommended postnatal checks?'; and (3) 'In 2019, out of 100 newborns, how many were receiving all the recommended vaccines on time?'. Both Panel B and C plot the treatment effect on the change in the average perceived program's effectiveness for workers depending on their prior perceived effectiveness levels. The first bar in each panel reports the treatment effect for those CHWs holding prior beliefs above the median, while the second bar shows the treatment effect for those holding prior beliefs below or equal to the median. All specifications include CHW-level strata fixed effects. They also include CHW-level controls. The full list of controls is presented in Section 4. Standard errors are clustered at the CHW level.

Table 6: Motivation measures

	CHWs face-to-face survey			
	Mission	Self-image	Identity	Warm-glow
	(1)	(2)	(3)	(4)
Treat	0.048	0.017	-0.043	-0.021
	(0.087)	(0.116)	(0.119)	(0.143)
	[1]	[1]	[1]	[1]
Observations	241	244	243	245
Control Mean	0.001	0.005	0.000	0.000
R^2	0.292	0.305	0.197	0.304

Note: In this table, I report the intent-to-treat estimates of the direct effects of the treatment on mission motivation measures. These are obtained by ordinary least squares (OLS) estimation of Equation 1. The estimation sample consists of all the CHWs interviewed in the post-intervention face-to-face survey. In all columns, the dependent variables are indexes built using the procedure in Kling, Liebman, and Katz (2007). I calculate within-sample z-scores for each individual outcome, employing the mean and the standard deviation of the pure control group. I then obtain the unweighted average z-score for each group of outcomes. Each column measures a different motivational dimension. Column (1) measures agreement with the following statements: 'You enjoy the CHWs program more than other volunteer activities because of the importance it places on the mission of improving health in your community.'; 'You believe that the mission of the CHWs program is very similar to your thinking from the time you joined the program.'; and 'If the mission of the CHWs program were different, you would not have been so attached to the program.' Column (2) measures agreement with the following statements: 'Your job 'fits' how you see yourself in the future.' and 'Your job 'fits' how you see yourself.' Column (3) measures agreement with the following statements: 'You like the identity your job gives you.' and 'The work you do on this job helps you satisfy who you are.' Column (4) measures agreement with the following statement: 'Supporting other people makes you very happy.' All specifications include CHW-level strata fixed effects. They also include CHW-level controls. The full list of controls is presented in Section 4. Standard errors are reported in parentheses. Standard errors are clustered at the CHW level. *** p<0.01, ** p<0.05, * p<0.1.

References

Abeler, Johannes, Armin Falk, Lorenz Goette, and David Huffman. 2011. "Reference Points and Effort Provision." *American Economic Review* 101 (2): 470–492.

Akerlof, George A, and Rachel E Kranton. 2000. "Economics and Identity." *Quarterly Journal of Economics,* 115 (3): 715–753.

———. 2005. "Identity and the Economics of Organizations." *Journal of Economic Perspectives,* 19 (1): 9–32.

Anderson, Ellen, Travis J Lybbert, Ashish Shenoy, Rupika Singh, and Daniel Stein. 2023. "Does Survey Mode Matter? Comparing In-Person and Phone Agricultural Surveys in India." *Mimeo* .

Anderson, Michael L. 2008. "Multiple Inference and Gender Differences in the Effects of Early Intervention: A Reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects." *Journal of the American statistical Association* 103 (484): 1481–1495.

Andreoni, James. 1990. "Impure Altruism and Donations to Public Goods: A Theory of Warm-glow Giving." *The Economic Journal* 100 (401): 464–477.

Ashraf, Nava, and Oriana Bandiera. 2018. "Social Incentives in Organizations." *Annual Review of Economics* 10: 439–463.

Ashraf, Nava, Oriana Bandiera, Edward Davenport, and Scott S Lee. 2020. "Losing Prosociality in the Quest for Talent? Sorting, Selection, and Productivity in the Delivery of Public Services." *American Economic Review* 110 (5): 1355–94.

Ashraf, Nava, Oriana Bandiera, and B Kelsey Jack. 2014. "No Margin, no Mission? A Field Experiment on Incentives for Public Service Delivery." *Journal of Public Economics,* 120: 1–17.

Ashraf, Nava, Oriana Bandiera, and Scott S Lee. 2014. "Awards Unbundled: Evidence from a Natural Field Experiment." *Journal of Economic Behavior & Organization,* 100: 44–63.

Bandiera, Oriana, Robin Burgess, Erika Deserranno, Ricardo Morel, Munshi Sulaiman, and Imran Rasul. 2023. "Social Incentives, Delivery Agents, and the Effectiveness of Development Interventions." *Journal of Political Economy Microeconomics* 1 (1): 162–224.

Banuri, Sheheryar, Stefan Dercon, and Varun Gauri. 2019. "Biased Policy Professionals." *The World Bank Economic Review* 33 (2): 310–327.

Barnett, Michael L, Irene Henriques, and Bryan W Husted. 2020. "Beyond Good Intentions: Designing CSR Initiatives for Greater Social Impact." *Journal of Management* 46 (6): 937–964.

Bénabou, Roland. 2013. "Groupthink: Collective Delusions in Organizations and Markets." *Review of Economic Studies* 80 (2): 429–462.

Bénabou, Roland, and Jean Tirole. 2003. "Intrinsic and Extrinsic Motivation." *Review of Economic Studies,* 70 (3): 489–520.

———. 2006. "Incentives and Prosocial Behavior." *American Economic Review,* 96 (5): 1652–1678.

———. 2010. "Individual and Corporate Social Responsibility." *Economica* 77 (305): 1–19.

———. 2016. "Mindful Economics: The production, Consumption, and Value of Beliefs." *Journal of Economic Perspectives* 30 (3): 141–164.

Benjamini, Yoav, Abba M Krieger, and Daniel Yekutieli. 2006. "Adaptive Linear Step-up Procedures that Control the False Discovery Rate." *Biometrika* 93 (3): 491–507.

Besley, Timothy, Robin Burgess, Adnan Khan, and Guo Xu. 2022. "Bureaucracy and Development." *Annual Review of Economics* 14: 397–424.

Besley, Timothy, and Maitreesh Ghatak. 2005. "Competition and Incentives with Motivated Agents." *American Economic Review,* 95 (3): 616–636.

———. 2018. "Prosocial Motivation and Incentives." *Annual Review of Economics* 10: 411–438.

Björkman Nyqvist, Martina, Andrea Guariso, Jakob Svensson, and David 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.

Bloom, Nicholas, Benn Eifert, Aprajit Mahajan, David McKenzie, and John Roberts. 2013. "Does Management Matter? Evidence from India." *The Quarterly Journal of Economics* 128 (1): 1–51.

Bloom, Nicholas, Renata Lemos, Raffaella Sadun, and John Van Reenen. 2015. "Does Management Matter in Schools?" *The Economic Journal* 125 (584): 647–674.

Bloom, Nicholas, and John Van Reenen. 2010. "Why do Management Practices Differ across Firms and Countries?" *Journal of Economic Perspectives* 24 (1): 203–224.

Camerer, Colin, Linda Babcock, George Loewenstein, and Richard Thaler. 1997. "Labor Supply of New York City Cabdrivers: One Day at a Time." *The Quarterly Journal of Economics* 112 (2): 407–441.

Casey, Katherine, Rachel Glennerster, and Edward Miguel. 2012. "Reshaping Institutions: Evidence on Aid Impacts Using a Preanalysis Plan." *The Quarterly Journal of Economics* 127 (4): 1755–1812.

Cassar, Lea, and Jesper Armouti-Hansen. 2020. "Optimal Contracting with Endogenous Project Mission." *Journal of the European Economic Association* 18 (5): 2647–2676.

Cassar, Lea, and Stephan Meier. 2018. "Nonmonetary Incentives and the Implications of Work as a Source of Meaning." *Journal of Economic Perspectives* 32 (3): 215–238.

Chaudhury, Nazmul, Jeffrey Hammer, Michael Kremer, Karthik Muralidharan, and F Halsey Rogers. 2006. "Missing in Action: Teacher and Health Worker Absence in Developing Countries." *Journal of Economic Perspectives* 20 (1): 91–116.

Chou, Yuan K. 2002. "Testing Alternative Models of Labour Supply: Evidence from Taxi Drivers in Singapore." *The Singapore Economic Review* 47 (01): 17–47.

Christopher, Jason B, Alex Le May, Simon Lewin, and David A Ross. 2011. "Thirty Years After Alma-Ata: a Systematic Review of the Impact of Community Health Workers Delivering Curative Interventions Against Malaria, Pneumonia and Diarrhoea on Child Mortality and Morbidity in Sub-Saharan Africa." *Human Resources for Health,* : 9–27.

Crawford, Vincent P, and Juanjuan Meng. 2011. "New York City Cab Drivers' Labor Supply Revisited: Reference-dependent Preferences with Rational-expectations Targets for Hours and Income." *American Economic Review* 101 (5): 1912–1932.

Dal Bó, Ernesto, Frederico Finan, and Martín A Rossi. 2013. "Strengthening State Capabilities: The Role of Financial Incentives in the Call to Public Service." *The Quarterly Journal of Economics* 128 (3): 1169–1218.

Dana, Jason, Roberto A Weber, and Jason Xi Kuang. 2007. "Exploiting Moral Wiggle Room: Experiments Demonstrating an Illusory Preference for Fairness." *Economic Theory,* 33 (1): 67–80.

Delavande, Adeline, Xavier Giné, and David McKenzie. 2011. "Measuring Subjective Expectations in Developing Countries: A Critical Review and New Evidence." *Journal of Development Economics* 94 (2): 151–163.

Delfgaauw, Josse, and Robert Dur. 2008. "Incentives and Workers' Motivation in the Public Sector." *The Economic Journal* 118 (525): 171–191.

Deserranno, Erika. 2019. "Financial Incentives as Signals: Experimental Evidence from the Recruitment of Village Promoters in Uganda." *American Economic Journal: Applied Economics,* 11 (1): 277–317.

Deserranno, Erika, Stefano Caria, Philipp Kastrau, and Gianmarco León-Ciliotta. 2020. "Financial Incentives in Multilayered Organizations: Empirical Evidence from the Community Health Worker Program in Sierra Leone." .

Deserranno, Erika, Philipp Kastrau, and Gianmarco León Ciliotta. 2021. "Promotions and Productivity: The Role of Meritocracy and Pay Progression in the Public Sector." *CEPR Discussion Paper* .

Dhaliwal, Iqbal, Esther Duflo, Rachel Glennerster, and Caitlin Tulloch. 2013. "Comparative Cost-effectiveness Analysis to Inform Policy in Developing Countries: a General Framework with Applications for Education." *Education Policy in Developing Countries* 17: 285–338.

Dixit, Avinash. 2002. "Incentives and Organizations in the Public Sector: An Interpretative Review." *Journal of Human Resources* : 696–727.

Doran, Kirk. 2014. "Are Long-term Wage Elasticities of Labor Supply More Negative than Short-term Ones?" *Economics Letters* 122 (2):208–210.

Duflo, Esther, Rema Hanna, and Stephen P Ryan. 2012. "Incentives Work: Getting Teachers to Come to School." *American Economic Review* 102 (4): 1241–78.

Dunning, Thad, Guy Grossman, Macartan Humphreys, Susan D Hyde, Craig McIntosh, and Gareth Nellis. 2019a. *Information, Accountability, and Cumulative Learning: Lessons from Metaketa I.* Cambridge University Press.

Dunning, Thad, Guy Grossman, Macartan Humphreys, Susan D Hyde, Craig McIntosh, Gareth Nellis, Claire L Adida, Eric Arias, Clara Bicalho, Taylor C Boas et al. 2019b. "Voter Information Campaigns and Political Accountability: Cumulative Findings from a Preregistered Meta-analysis of Coordinated Trials." *Science Advances* 5 (7): eaaw2612.

Eil, David, and Justin M Rao. 2011. "The Good News-bad News Effect: Asymmetric Processing of Objective Information about Yourself." *American Economic Journal: Microeconomics* 3 (2): 114–138.

Ellingsen, Tore, and Magnus Johannesson. 2008. "Pride and Prejudice: The Human Side of Incentive Theory." *American Economic Review* 98 (3): 990–1008.

Exley, Christine L. 2020. "Using Charity Performance Metrics as an Excuse not to Give." *Management Science* 66 (2): 553–563.

Exley, Christine L, and Judd B Kessler. 2021. "Information Avoidance and Image Concerns." Tech. rep., National Bureau of Economic Research.

Farber, Henry S. 2005. "Is Tomorrow Another Day? The Labor Supply of New York City Cabdrivers." *Journal of Political Economy* 113 (1): 46–82.

———. 2008. "Reference-dependent Preferences and Labor Supply: The Case of New York City Taxi Drivers." *American Economic Review* 98 (3): 1069–1082.

———. 2015. "Why You Can't Find a Taxi in the Rain and Other Labor Supply Lessons from Cab Drivers." *The Quarterly Journal of Economics* 130 (4): 1975–2026.

Fenizia, Alessandra. 2022. "Managers and Productivity in the Public Sector." *Econometrica* 90 (3): 1063–1084.

Finan, Frederico, Benjamin A Olken, and Rohini Pande. 2017. "The Personnel Economics of the Developing State." *Handbook of Economic Field Experiments* 2: 467–514.

Fioretti, Michele. 2022. "Caring or Pretending to Care? Social Impact, Firms' Objectives, and Welfare." *Journal of Political Economy* 130 (11): 2898–2942.

Fracchia, Mattia, Teresa Molina-Millán, and Pedro C Vicente. 2023. "Motivating Volunteer Health Workers in an African Capital City." *Journal of Development Economics* : 103096.

Francois, Patrick. 2000. "'Public Service Motivation' as an Argument for Government Provision." *Journal of Public Economics* 78 (3): 275–299.

Frey, Bruno S, and Reto Jegen. 2001. "Motivation Crowding Theory." *Journal of Economic Surveys* 15 (5): 589–611.

Gertner, Robert H. 2023. "The Organization of Social Enterprises." *Annual Review of Economics* 15.

Gillan, Stuart L, Andrew Koch, and Laura T Starks. 2021. "Firms and Social Responsibility: A Review of ESG and CSR Research in Corporate Finance." *Journal of Corporate Finance* 66: 101889.

Gilmore, Brynne, and Eilish McAuliffe. 2013. "Effectiveness of Community Health Workers Delivering Preventive Interventions for Maternal and Child Health in Low- and Middle-income Countries: a Systematic Review." *BMC Public Health,* 13 (1): 847.

Giné, Xavier, Ghazala Mansuri, and Slesh A Shrestha. 2022. "Mission and the Bottom Line: Performance Incentives in a Multigoal Organization." *Review of Economics and Statistics* 104 (4): 748–763.

Glewwe, Paul, Nauman Ilias, and Michael Kremer. 2010. "Teacher Incentives." *American Economic Journal: Applied Economics* 2 (3): 205–27.

Gneezy, Uri, Stephan Meier, and Pedro Rey-Biel. 2011. "When and Why Incentives (Don't) Work to Modify Behavior." *Journal of Economic Perspectives* 25 (4): 191–210.

Golman, Russell, David Hagmann, and George Loewenstein. 2017. "Information Avoidance." *Journal of Economic Literature* 55 (1): 96–135.

Grant, Adam M. 2008. "Does Intrinsic Motivation Fuel the Prosocial Fire? Motivational Synergy in Predicting Persistence, Performance, and Productivity." *Journal of Applied Psychology* 93 (1): 48.

Grossman, Zachary, and Joel J Van Der Weele. 2017. "Self-image and willful ignorance in social decisions." *Journal of the European Economic Association* 15 (1): 173–217.

Hackman, J Richard, and Greg R Oldham. 1975. "Development of the Job Diagnostic Survey." *Journal of Applied Psychology* 60 (2): 159.

Haines, Andy, David Sanders, Uta Lehmann, Alexander K Rowe, Joy E Lawn, Steve Jan, Damian G Walker, and Zulfiqar Bhutta. 2007. "Achieving Child Survival Goals: Potential Contribution of Community Health Workers." *The Lancet,* 369 (9579): 2121–2131.

Hjort, Jonas, Diana Moreira, Gautam Rao, and Juan Francisco Santini. 2021. "How Research Affects Policy: Experimental Evidence from 2,150 Brazilian Municipalities." *American Economic Review* 111 (5): 1442–80.

Huffman, David, Collin Raymond, and Julia Shvets. 2022. "Persistent Overconfidence and Biased Memory: Evidence from Managers." *American Economic Review* 112 (10): 3141–3175.

Ioannidis, John PA. 2005. "Why Most Published Research Findings are False." *PLoS medicine* 2 (8): e124.

Jensen, Robert. 2010. "The (Perceived) Returns to Education and the Demand for Schooling." *The Quarterly Journal of Economics* 125 (2): 515–548.

Kamenica, Emir. 2012. "Behavioral economics and psychology of incentives." *Annual Review of Economics* 4 (1): 427–452.

Karlan, Dean, and Daniel H Wood. 2017. "The Effect of Effectiveness: Donor Response to Aid Effectiveness in a Direct Mail Fundraising Experiment." *Journal of Behavioral and Experimental Economics* 66: 1–8.

Kaufmann, Katja Maria. 2014. "Understanding the Income Gradient in College Attendance in Mexico: The Role of Heterogeneity in Expected Returns." *Quantitative Economics* 5 (3): 583–630.

Khan, Adnan Q, Asim I Khwaja, and Benjamin A Olken. 2016. "Tax Farming Redux: Experimental Evidence on Performance Pay for Tax Collectors." *The Quarterly Journal of Economics* 131 (1): 219–271.

Khan, Muhammad Yasir. 2020. "Mission Motivation and Public Sector Performance: Experimental Evidence from Pakistan." *Mimeo* .

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

Kőszegi, Botond, and Matthew Rabin. 2006. "A model of Reference-dependent Preferences." *The Quarterly Journal of Economics* 121 (4): 1133–1165.

Malmendier, Ulrike, and Geoffrey Tate. 2005. "CEO Overconfidence and Corporate Investment." *The Journal of Finance* 60 (6): 2661–2700.

McKenzie, David. 2012. "Beyond Baseline and Follow-up: The Case for More T in Experiments." *Journal of Development Economics,* 99 (2): 210–221.

Metzger, Laura, and Isabel Günther. 2019a. "Is It What You Say or How You Say It? The Impact of Aid Effectiveness Information and Its Framing on Donation Behavior." *Journal of Behavioral and Experimental Economics* 83: 101461.

———. 2019b. "Making an Impact? The Relevance of Information on Aid Effectiveness for Charitable Giving. A Laboratory Experiment." *Journal of Development Economics* 136: 18–33.

Möbius, Markus M, Muriel Niederle, Paul Niehaus, and Tanya S Rosenblat. 2022. "Managing Self-confidence: Theory and Experimental Evidence." *Management Science* 68 (11): 7793–7817.

Murdock, Kevin. 2002. "Intrinsic Motivation and Optimal Incentive Contracts." *RAND Journal of Economics* : 650–671.

Nguyen, Trang. 2008. "Information, Role Models and Perceived Returns to Education: Experimental Evidence from Madagascar." *Unpublished manuscript* 6.

Null, Clair. 2011. "Warm Glow, Information, and Inefficient Charitable Giving." *Journal of Public Economics* 95 (5-6):455–465.

Our World in Data team. 2023. "Ensure Healthy Lives and Promote Well-being for All at All Ages." Tech. rep. Published online at OurWorldInData.org. Retrieved from: 'https://ourworldindata.org/sdgs/good-health-wellbeing' [Online Resource].

Rasul, Imran, and Daniel Rogger. 2018. "Management of Bureaucrats and Public Service Delivery: Evidence from the Nigerian Civil Service." *The Economic Journal* 128 (608): 413–446.

Rogger, Daniel and Ravi Somani. 2023. "Hierarchy and information." *Journal of Public Economics* 219:104823.

Rotemberg, Julio J. 1994. "Human Relations in the Workplace." *Journal of Political Economy* 102 (4): 684–717.

Russo, Giuliano, Enrico Pavignani, Catia Sá Guerreiro, and Clotilde Neves. 2017. "Can We Halt Health Workforce Deterioration in Failed States? Insights from Guinea-Bissau on the Nature, Persistence and Evolution of its HRH Crisis." *Human Resources for Health* 15 (1): 1–14.

Steele, Claude M. 1988. "The Psychology of Self-affirmation: Sustaining the Integrity of the Self." In *Advances in Experimental Social Psychology*, vol. 21. Elsevier, 261–302.

Taylor, Celia, Frances Griffiths, and Richard Lilford. 2017. "Affordability of Comprehensive Community Health Worker Programmes in Rural Sub-Saharan Africa." *BMJ Glob Health,* 2 (3): e000391.

Thakral, Neil, and Linh T Tô. 2021. "Daily Labor Supply and Adaptive Reference Points." *American Economic Review* 111 (8): 2417–2443.

Vivalt, Eva. 2020. "How Much Can We Generalize from Impact Evaluations?" *Journal of the European Economic Association* 18 (6): 3045–3089.

Vivalt, Eva, and Aidan Coville. 2023. "How do Policymakers Update their Beliefs?" *Journal of Development Economics* : 103121.

Williams, Martin J. 2020. "External Validity and Policy Adaptation: From Impact Evaluation to Policy Design." *The World Bank Research Observer* 35 (2):158–191.

World Bank. 2019. "Guinea-Bissau: Service Delivery Indicators Report - Health." Available at https://openknowledge.worldbank.org/handle/10986/32029.

———. 2020. "World Development Indicators." Available at https://data.worldbank.org/indicator/NY.GDP.PCAP.CD.

———. 2021. "World Development Report 2021: Data for Better Lives." Available at https://www.worldbank.org/en/publication/wdr2021.

———. 2023. "World Development Indicators." Available at https://databank.worldbank.org/source/world-development-indicators.

Zambruni, Jerome Pfaffmann, Kumanan Rasanathan, David Hipgrave, Nathan P Miller, Maureen Momanyi, Luwei Pearson, Dolores Rio, Magali Romedenne, Sagri Singh, Mark Young et al. 2017. "Community Health Systems: Allowing Community Health Workers to Emerge from the Shadows." *The Lancet Global Health,* 5 (9): e866–e867.

Zimmermann, Florian. 2020. "The Dynamics of Motivated Beliefs." *American Economic Review* 110 (2): 337–363.

ONLINE APPENDIX

Does Performance Evidence Motivate?

A Field Experiment in Guinea-Bissau's Health Sector.

Mattia Fracchia

A CHW program details

- **Perinatal care**:
    1. Prenatal checks for the pregnant woman
    2. Post-natal care for the newborn and the mother
- **Child care**:
    1. Breastfeeding
    2. Child malnutrition
    3. Child vaccination
- **Diarrhea, Malaria, Pneumonia**:
    1. Diarrhea prevention and treatment
    2. Malaria prevention and treatment
    3. Pneumonia prevention and treatment
- **HIV, TBC**:
    1. HIV prevention
    2. Tuberculosis prevention and searching
- **Preventive practices at home**:
    1. Impregnated mosquito nets use
    2. Hand washing
    3. Water treatment with bleach

B Timeline

Figure B1 presents the timeline of the project. It is worth noting that the break between the intervention and the follow-up phone survey corresponds to the rainy season in the country, i.e., June to October.

Figure B1: Project Timeline

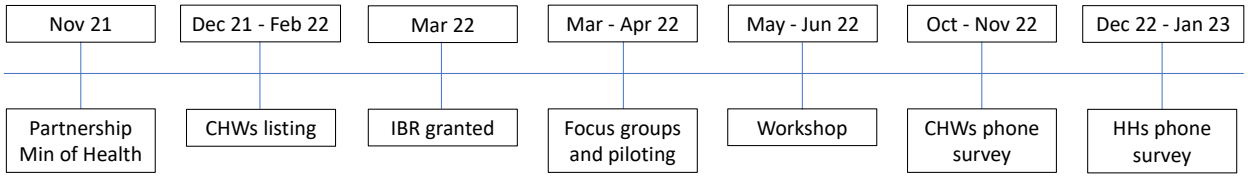
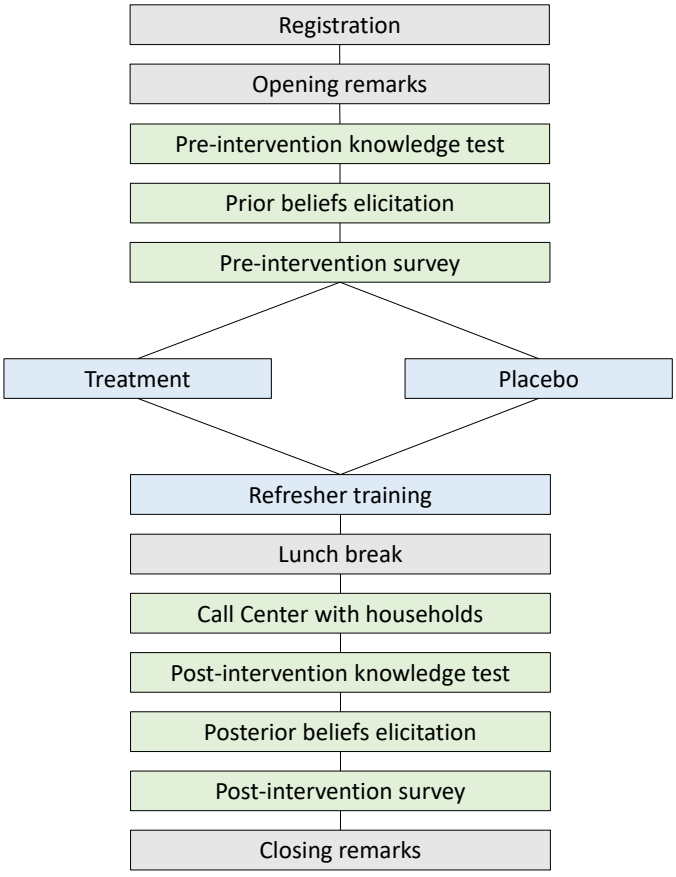


Figure B2 presents the structure of the workshop. The green boxes represent data collection activities facilitated by the team's enumerators, while the blue boxes represent information delivery activities facilitated by the project's trainer. To limit concerns of contamination, a workshop could be either treatment or control, i.e., only one between 'Treatment' and 'Placebo' was delivered. The workshops took all day and were held in the same location where CHWs generally meet for the program's monthly meeting. The head of the local health center and the local program supervisor were always invited to attend the workshop's opening remarks but did not stay for the rest of the activities.

Figure B2: Workshop structure



C Intervention

Figure C1: Intervention delivery



C.1 Anecdotal evidence

Below is a summary of the script of the anecdotal evidence introduction.

- Once upon a time, in a peaceful village near a winding river in the Oio and Farim region, lived Binta and Mário. They fell in love amidst the cashew trees during the 2013 campaign, culminating in a joyful wedding celebration. In October 2014, at the end of the rainy season, they welcomed their first daughter, Justina.

- Unfortunately, soon, Justina fell ill. Fearing for their child, Binta and Mário turned to the village's wise matriarchs for guidance. They suggested that Justina's symptoms were due to living near the river. However, despite their years of experience, they could agree on the correct diagnosis or treatment. Binta and Mário subsequently consulted with the traditional healer, who said the child was afflicted by an "irã," an evil spirit. The healer suspected that the devil had impregnated Binta because of her habit of bathing in the late afternoon, a time when evil entities were thought to be most active. Dismayed by the traditional healer's diagnosis, Binta and Mário were undecided about what to do. They thought about going to the health center, but it took them a while to decide. They waited a few more days to see if Justina would get better. Unfortunately, the child got worse and tragically passed away in December 2014.

- Seasons shifted, and Binta and Mário's lives slowly returned to normal. A new cashew campaign arrived and passed. On the arrival of the 2015 rainy season, Binta found out she was pregnant again. In the spring of 2016, amid the vibrant festivities of Carnival, they welcomed their second daughter, Isabel.

- Later that year, a new community health program was announced. Some villagers embraced the opportunity to become Community Health Workers (CHWs). Some already performed similar functions but had no official recognition by the health system. Now, however, they had access to additional training and support from the program.

- In 2017, the program was officially launched, and CHWs began visiting households in the village. A well-prepared and dedicated CHW visits Binta, Mário, and Isabel. With kindness and determination, the CHW earned their trust. The CHW then begins to promote good health practices in the household. Binta likes the simple and respectful way the CHW uses to teach her these good

practices. During her visits, she learns many new things about the best way to ensure the health and well-being of Isabel and the whole family. Though their perspectives occasionally diverged, mutual respect and constructive dialogue bridged the gaps, gradually convincing Binta of the benefits of the CHW's knowledge. Eventually, she also gets CHW's telephone number. As months flowed by, Binta, Mário, and Isabel flourished as a family, cherished by their village.

- Unfortunately, in 2019, adversity struck once more when Isabel fell ill. As in Justina's case, Binta and Mário consulted the village's wise matriarchs and the traditional healer for advice. As in Justina's case, they suggest different possible diagnoses and treatments. All the alternatives put forward will take some time to prepare and implement.

- This time, though, they turned to their trusted CHW for guidance. Binta called the CHW, who visited her house the same day. The CHW recognized the gravity of Isabel's condition and insisted on immediate evacuation to the health center. Mindful of their past experiences, Binta and Mário accepted the CHW's advice. The CHW swiftly coordinated Isabel's journey to the health center, ensuring she received timely care.

- Fortune favored their decision, and Isabel regained her health and strength in a few days. The village's traditional healer and wise women rejoiced at the quick recovery, extending their gratitude to the CHW, now acknowledged as the village's health caretaker. Binta and Mário expressed their profound appreciation for the CHW's support.

- From that pivotal moment onward, Binta and Mário turned to the CHW whenever health concerns arose. Today, in 2022, Isabel continues to thrive, and she also has a little brother, Domingos.

C.2 Scientific evidence

Figure C2: Under-5 mortality rate



Figure C3: Share of pregnant women who completed at least four prenatal checks

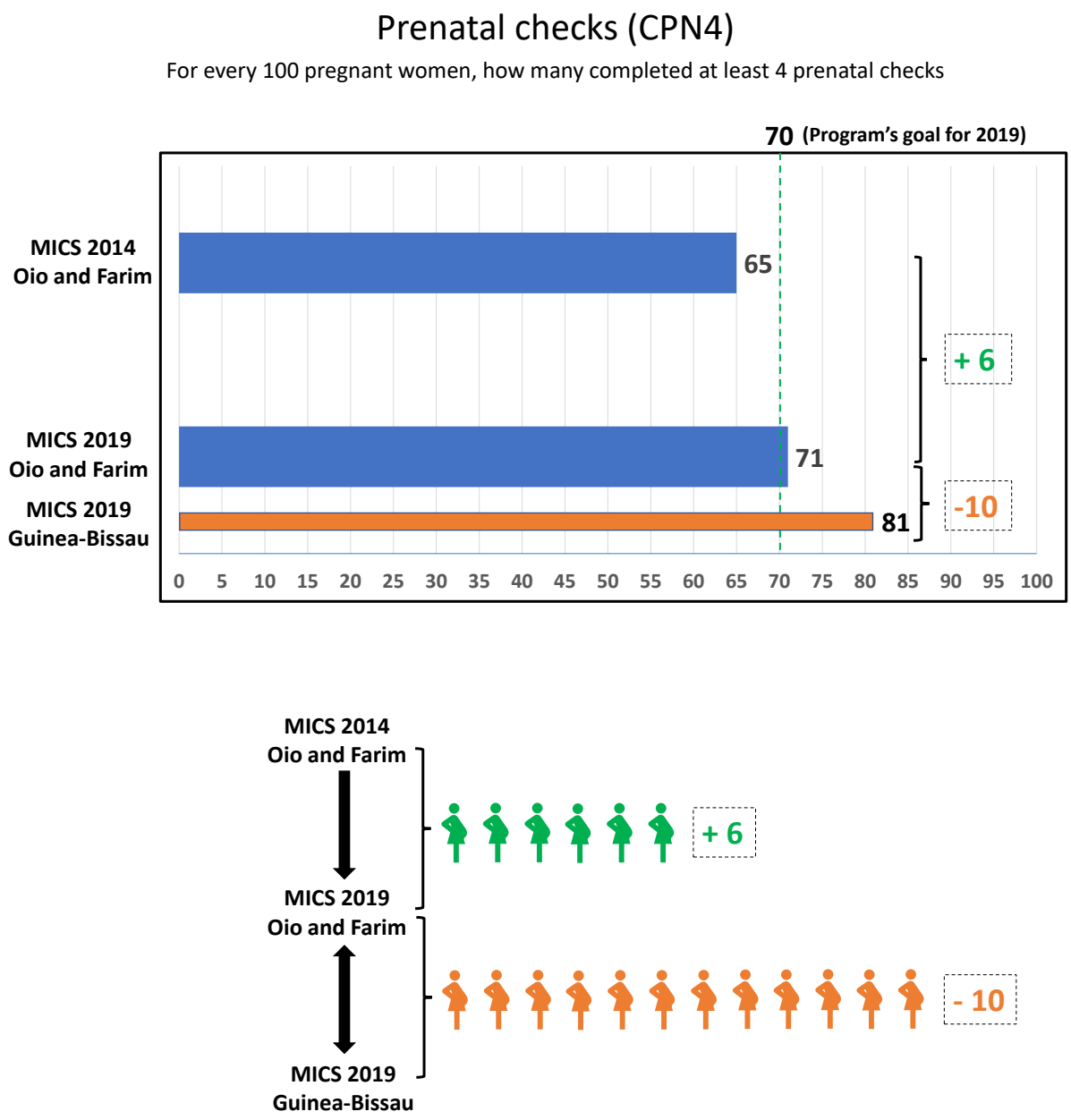
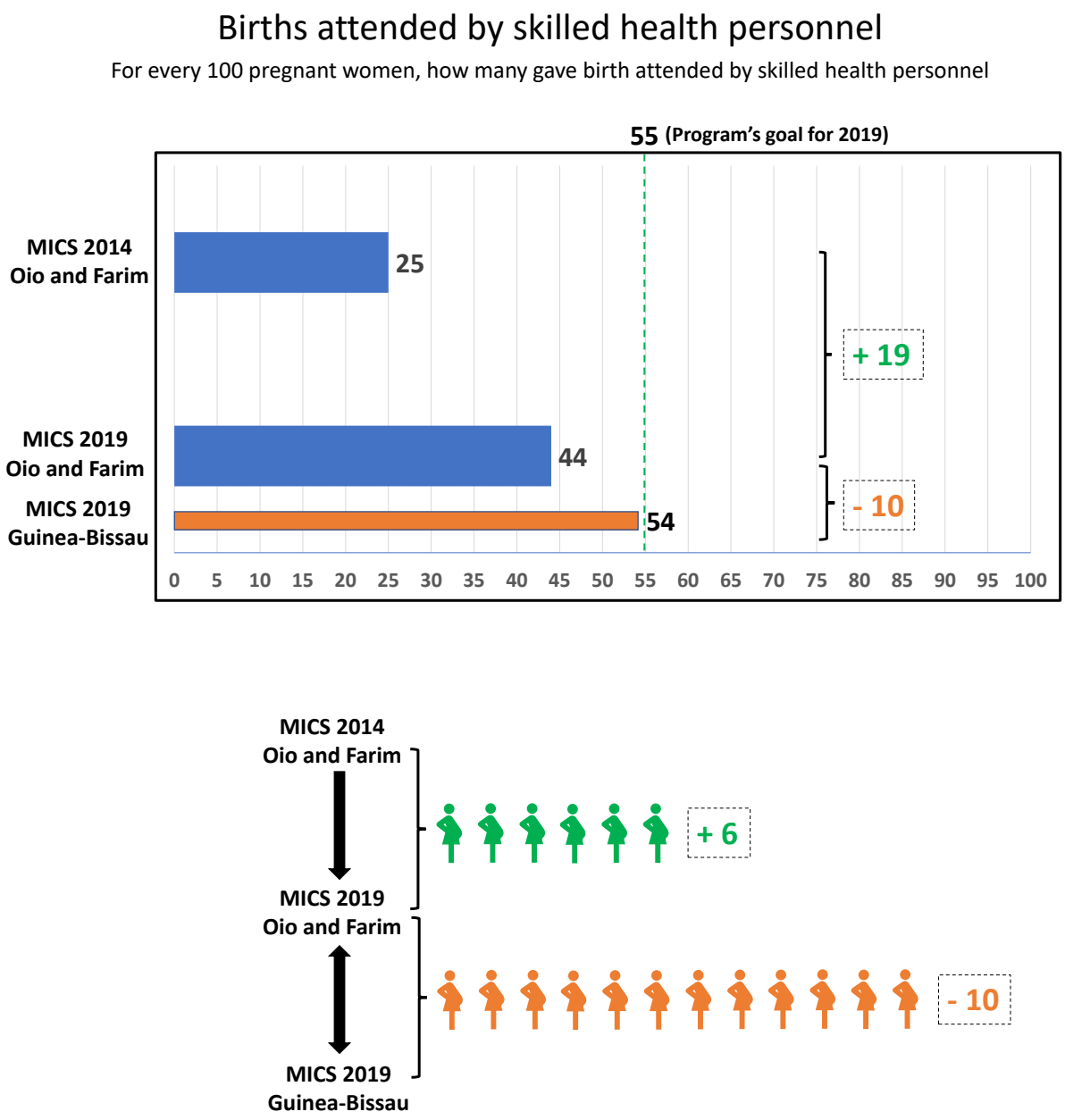


Figure C4: Share of births attended by skilled health personnel



D Randomization

Figure D1 shows assignments to treatment groups of participants. Figure D2 shows household sampling.

Figure D1: Experimental design (CHWs) - CONSORT diagram

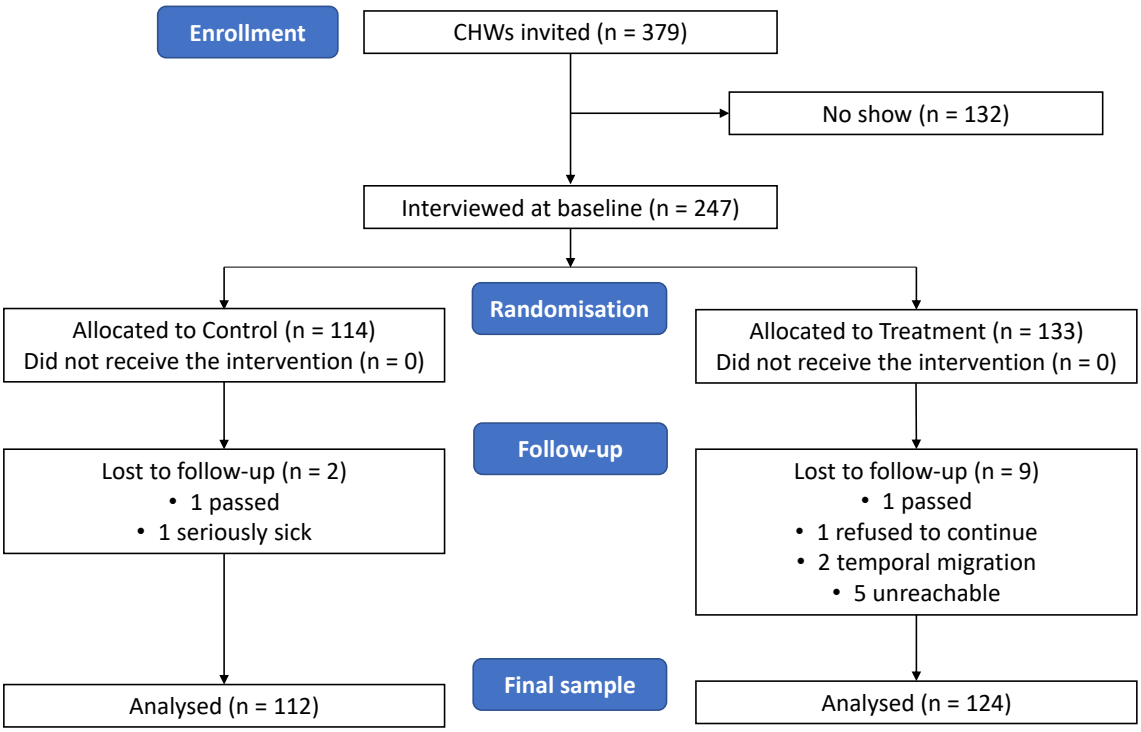


Figure D2: Experimental design (HHs) - CONSORT diagram

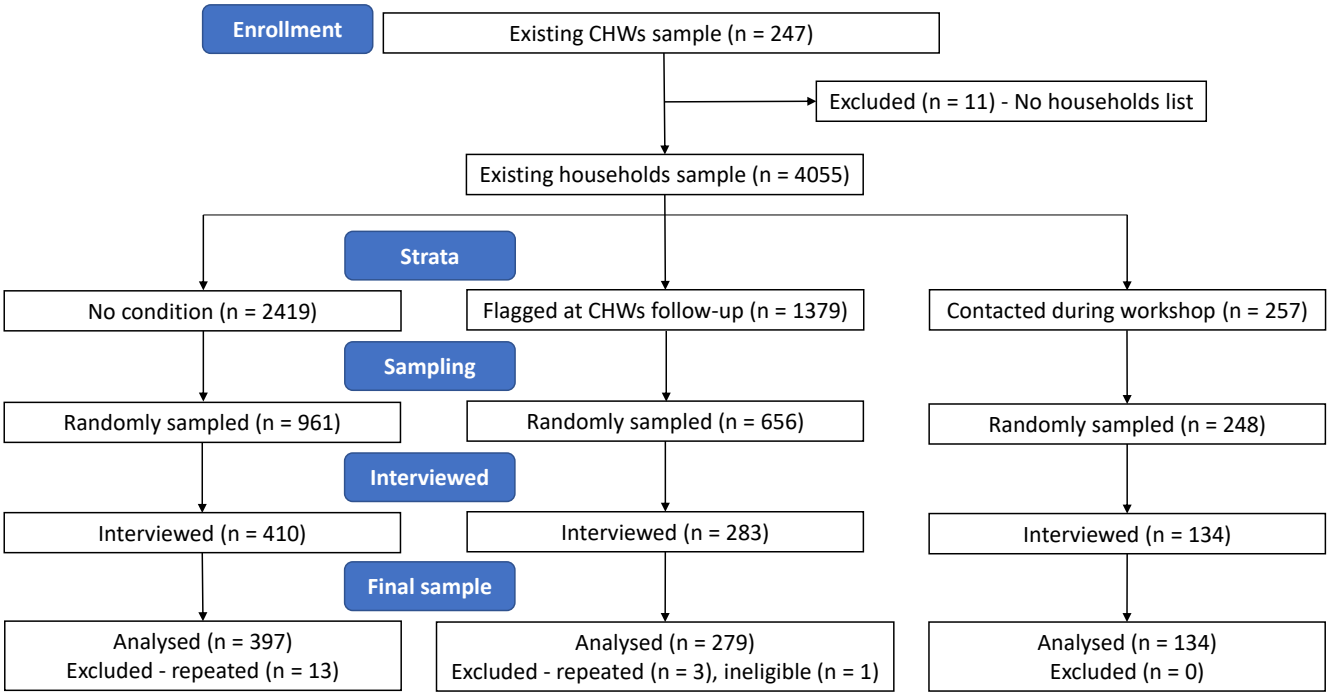


Table D1: Minimum detectable effect sizes for outcomes presented in Table 1

	HHs phone survey						
	Past month			Past six months			
	Visits received	Visit duration (minutes)	Time with the CHW (minutes)	Visits received	Visit duration (minutes)	Time with the CHW (minutes)	Satisfaction with the program
	(1)	(2)	(3)	(4)	(5)	(6)	
Treatment vs. Control	0.134**	0.206**	0.275***	0.004	0.177**	0.140**	0.145*
	(0.068)	(0.082)	(0.084)	(0.076)	(0.071)	(0.069)	(0.079)
Ex-ante 2-sided no controls	0.20	0.20	0.20	0.20	0.20	0.20	0.20
Ex-ante 1-sided no controls	0.18	0.18	0.18	0.18	0.18	0.18	0.18
Ex-ante 2-sided controls	0.17	0.20	0.17	0.21	0.17	0.16	0.22
Ex-ante 1-sided controls	0.15	0.18	0.15	0.18	0.15	0.14	0.19
Ex-post 2-sided no controls	0.23	0.23	0.27	0.20	0.24	0.24	0.20
Ex-post 1-sided no controls	0.11	0.11	0.14	0.10	0.12	0.12	0.10
Ex-post 2-sided controls	0.19	0.23	0.24	0.21	0.20	0.19	0.22
Ex-post 1-sided controls	0.10	0.11	0.12	0.11	0.10	0.10	0.11

Note: Estimates report the Minimum Detectable Effect Sizes (MDES) for the outcomes presented in Table 1, in standard deviations, assuming a power level of 0.8 and a significance level of 0.05. For reference, I report the treatment effects in standard deviations in the first row. For each outcome, I report eight MDES. Ex-ante 2-sided no controls MDES is computed using the Stata command *power* considering a control group size of 368 and a treatment group size of 442. Ex-ante 1-sided no control MDES is computed using the Stata command *power* considering the same group sizes as Ex-ante 2-sided no control, and using the option *onesided*, assuming a positive treatment effect. Ex-ante 2-sided controls MDES is computed by multiplying the coefficient of Ex-ante 2-sided no controls by the ratio of the standard error of the beta coefficient of the treatment dummy in the OLS regression with controls over the standard error of the beta coefficient of the treatment dummy in the OLS regression without controls. Ex-ante 1-sided control MDES is computed by multiplying the coefficient of Ex-ante 1-sided no control by the ratio of the standard error of the beta coefficient of the treatment dummy in the OLS regression with controls over the standard error of the beta coefficient of the treatment dummy in the OLS regression without controls. Ex-post 2-sided no controls MDES is computed by multiplying the standard error of the beta coefficient of the treatment dummy in the OLS regression without controls by 2.80. Ex-post 1-sided no controls MDES is computed by multiplying the standard error of the beta coefficient of the treatment dummy in the OLS regression without controls by 1.40, assuming a positive treatment effect. Ex-post 2-sided controls MDES is computed by multiplying the standard error of the beta coefficient of the treatment dummy in the OLS regression with controls by 2.80. Ex-post 1-sided controls MDES is computed by multiplying the standard error of the beta coefficient of the treatment dummy in the OLS regression with controls by 1.40, assuming a positive treatment effect.

E Additional details on measurement

E.1 Beliefs elicitation

Beliefs about the status of health indicators at the local level

I elicit CHWs' beliefs about the level of a set of the CHWs program's target health indicators. The enumerators prompted respondents to answer all questions using a one-meter ruler modified to serve as a slider with a 100 scale - see Figure E1.

Enumerators used the following message to explain how to use the slider: "*I will ask several questions about how many people within a group of 100 will perform a certain action. Choose 0 if your answer is "no one", choose 1 if your answer is "only 1", and so on, up to 100, which reflects the answer "all". Let me give you an example. If I ask you, "Out of 100 people in Bissau, how many do you think to own a cell phone?" If you answer 87, that means you think that, on average, out of every 100 people, 87 own a cell phone in Bissau.*"

To maximize comparability, enumerators specified the relevant sample and period: "*Let's consider the people who live in the Oio and Farim region. Let's think of 100 people as a sample of the most common people living in the villages of Oio and Farim. The most recent independent indicators we have available are from 2019. So I want you to refer your answers to the year 2019.*"

The questions covered the following six health indicators:

1. In 2019, out of 100 children, how many would not survive to age 5?
2. In 2019, out of 100 pregnant women, how many completed the four recommended prenatal checks before delivery?
3. In 2019, out of 100 pregnant women, how many were giving birth attended by qualified health-care personnel (doctors, nurses, or trained midwives)?
4. In 2019, out of 100 pregnant women, how many were giving birth at the health center?
5. In 2019, out of 100 new mothers, how many completed all recommended postnatal checks?
6. In 2019, out of 100 newborns, how many were receiving all the recommended vaccines on time?

Figure E1: Beliefs elicitation



Perceived program effectiveness in improving health indicators

Each question on health indicators was immediately followed by one on the program's perceived effectiveness in improving that health indicator at the local level. These questions used a 5-item Likert scale that takes the values 1 'Not at all,' 2 'A little,' 3 'Moderately,' 4 'A lot,' and 5 'Extremely'.

1. How much do you think the evolution of the [*relevant health indicator*] in the Oio and Farim region is attributable to the ASC program?

E.2 Pen-and-Paper knowledge test

CHWs sat through a pre and post-test on knowledge of health practices. The test contained 'True or False' and 'Single-choice' questions. Each correct answer counts for one point. The items' order is the same as the test. A mistake or an empty answer counts as zero. Table 4 reports treatment effects on four averages. 'Full test' reports the simple average of all 26 questions. 'Antenatal care' reports the simple average of questions 1–6 and 8–14. 'HIV prevention and testing' reports the simple average of questions 7 and 15–20. 'Danger signs in pregnant women' reports the simple average of questions 21–26.

1. The prenatal check aims to prepare the mother for motherhood. [True or False]
2. Only the first prenatal check is free. [True or False]
3. The prenatal check aims to prevent communicable diseases between the mother and her baby. [True or False]

4. The prenatal check is free only if the mother shows danger signs. [True or False]
5. The prenatal check aims to ensure a healthy pregnancy and diagnose possible complications during pregnancy. [True or False]
6. During the prenatal check, mothers receive the tetanus vaccine (VAT). [True or False]
7. In the prenatal check, mothers receive HIV testing. [True or False]
8. In the prenatal check, mothers receive the vaccine to prevent cholera. [True or False]
9. In the prenatal check, mothers receive prophylaxis for anemia and malaria. [True or False]
10. At the prenatal check, mothers receive a COVID-19 test. [True or False]
11. Select the correct statement:
    (a) Women should have their first prenatal check before they become pregnant.
    (b) Pregnant women should have their first prenatal check between the beginning of their pregnancy and 12/14 weeks (third month).
    (c) Pregnant women should have their first prenatal check between 12 weeks (third month) and 24 weeks (sixth month).
    (d) Pregnant women should have their first prenatal check after 24 weeks (the sixth month).
12. Select the correct statement:
    (a) A pregnant woman should have at least four prenatal checks at the health center in her first pregnancy alone.
    (b) A pregnant woman should have at least four prenatal checks at the health center each pregnancy.
    (c) A pregnant woman who has been a mother in the past needs to have only one prenatal check at the health center.
    (d) A pregnant woman should have more than one prenatal check at the health center only if the first one presents problems.
13. Select the correct statement:

(a) In the first pregnancy, a woman should receive a maximum of one tetanus vaccine (VAT) dose to protect herself and her baby.

(b) In the first pregnancy, a woman should receive two doses of tetanus vaccine (VAT) to protect herself and her baby.

(c) In the first pregnancy, a woman should receive one dose of tetanus vaccine (VAT for her protection and that of her baby. If you have twins, you should receive two doses.

(d) In the first pregnancy, the woman should receive five doses of tetanus vaccine (VAT).

14. Select the correct statement:

(a) Two doses of tetanus vaccine (VAT) are recommended to fully protect a woman during her reproductive life.

(b) Four doses of tetanus vaccine (VAT) are recommended to fully protect a woman during her reproductive life.

(c) Five doses of tetanus vaccine (VAT) are recommended to fully protect a woman during her reproductive life.

(d) Six doses of tetanus vaccine (VAT) are recommended to fully protect a woman during her reproductive life.

15. HIV stands for Human Immunodeficiency Virus. [True or False]

16. HIV can be transmitted through sweat or saliva. [True or False]

17. Mother-to-child transmission can occur during pregnancy, delivery, or breastfeeding in cases of pregnant women with uncontrolled HIV infection. [True or False]

18. An individual with HIV infection who is medicated and controlled (with an undetectable viral load) does not transmit the infection to others. [True or False]

19. All people with HIV will develop AIDS. [True or False]

20. Select the correct statement:

(a) It is advised that the pregnant woman and her husband or partner should be screened for HIV together.

(b) It is advised that the pregnant woman should be screened for HIV alone to respect her privacy.

(c) It is advised that the pregnant woman should be for HIV only in her first pregnancy.

(d) It is advised that the pregnant woman should be for HIV only if she shows danger signs.

21. Edemas in the feet, hands, and face are danger signs in pregnant women. [True or False]

22. Vaginal bleeding, loss of fluids with a bad smell, and pain in the lower abdomen are danger signs for pregnant women. [True or False]

23. It is a danger sign in pregnant women if the child does not move. [True or False]

24. Pregnant woman coughing for three days is a danger sign. [True or False]

25. Vomiting, fever, or severe headaches are danger signs for pregnant women. [True or False]

26. Select the correct statement:

    (a) A pregnant woman with danger signs should stay in bed and wait for it to pass.

    (b) A pregnant woman with danger signs should go immediately to the health center.

    (c) If the danger signs appeared in the first three months of pregnancy, there is no point in going to the health center.

    (d) If the pregnant woman has vaginal bleeding, it is because she is having her period.

E.3 Over-the-phone knowledge test

I compile a simplified version of the knowledge test to fit the time constraints of phone surveys. To identify 10 among the 26 questions of the pen-and-paper tests, I run a PCA for the control group's post-test grades and select the first 10 questions of the first PCA component. To adapt the questions to communication over the phone, I transformed single-choice questions into 'True or False' statements, i.e., questions 5 and 6 (*). Each correct answer counts for one point. A mistake or an empty answer counts as zero. Table I1 reports treatment effects on four averages. 'Full test' reports the simple average of all 10 questions. 'Antenatal care' reports the simple average of questions 1–3 and 5. 'HIV prevention and testing' reports the simple average of questions 4 and 6. 'Danger signs in pregnant women' reports the simple average of questions 7–10.

1. The prenatal check aims to prepare the mother for motherhood [True or False]
2. Only the first prenatal check is free [True or False]
3. The prenatal check aims to ensure a healthy pregnancy and diagnose possible complications during pregnancy [True or False]
4. In the prenatal check, mothers receive HIV testing [True or False]
5. A pregnant woman should have at least four prenatal checks at the health center in her first pregnancy alone [True or False]*
6. It is advised that the pregnant woman should be for HIV only if she shows danger signs [True or False]*
7. Edemas in the feet, hands, and face are danger signs in pregnant women [True or False]
8. Vaginal bleeding, loss of fluids with a bad smell, and pain in the lower abdomen are danger signs for pregnant women [True or False]
9. It is a danger sign in pregnant women if the child does not move [True or False]
10. Vomiting, fever, or severe headaches are danger signs for pregnant women [True or False]

F Balance

Table F1: Baseline balance on CHWs' characteristics

	CHWs face-to-face survey		
	N	Control Mean	Treatment
	(1)	(2)	(3)
Age	243	42.8	-0.26
			(1.52)
Gender (female =1)	247	0.035	-0.017
			(0.025)
Household head (=1)	245	0.75	0.0017
			(0.058)
Muslim (=1)	245	0.47	-0.023
			(0.051)
Education - 6 years or more	245	0.76	-0.050
			(0.059)
Currently studying (=1)	244	0.11	0.0017
			(0.043)
Paid job in the past 12 months	245	0.79	-0.16***
			(0.058)
Owns a business	245	0.35	0.083
			(0.069)
Number of villages assigned	247	1.96	0.056
			(0.15)
Number of households assigned	244	60.1	5.88
			(6.70)
Received initial training	247	0.97	0.0049
			(0.023)
Hours worked as CHW per week	242	12.5	-1.22
			(1.22)
Households visited in April 2022 (of assigned)	234	0.65	0.015
			(0.049)

Note. Column (1) reports the number of observations. Column (2) reports the sample mean of the control group. Column (3) reports estimates for the treatment indicator variable in equation (1). Standard errors, reported in parentheses, are clustered at the CHW level. *** p<0.01, ** p<0.05, * p<0.1.

Table F2: Baseline balance on CHWs' prior beliefs

	CHWs face-to-face survey		
	N	Control Mean	Treatment
	(1)	(2)	(3)
U5 mortality	243	4.36	-0.0083
			(0.89)
4 prenatal checks	243	14.4	1.53
			(1.54)
Assisted birth	243	28.9	0.14
			(2.06)
Birth at facility	243	35.4	-1.67
			(2.48)
Postnatal checks	243	30.4	2.99
			(2.27)
Complete vaccines	243	28.4	1.78
			(1.82)

Note. Column (1) reports the number of observations. Column (2) reports the sample mean of the control group. Column (3) reports estimates for the treatment indicator variable in equation (1). Standard errors, reported in parentheses, are clustered at the CHW level. *** p<0.01, ** p<0.05, * p<0.1.

Table F3: Baseline balance on CHWs' perceived program's contribution

	CHWs face-to-face survey		
	N	Control Mean	Treatment
	(1)	(2)	(3)
U5 mortality	242	3.96	0.046
			(0.10)
4 prenatal checks	243	4.01	-0.041
			(0.099)
Assisted birth	243	3.95	0.089
			(0.099)
Birth at facility	243	3.99	0.076
			(0.096)
Postnatal checks	243	3.95	0.12
			(0.097)
Complete vaccines	243	4.20	0.060
			(0.091)

Note. Column (1) reports the number of observations. Column (2) reports the sample mean of the control group. Column (3) reports estimates for the treatment indicator variable in equation (1). Standard errors, reported in parentheses, are clustered at the CHW level. *** p<0.01, ** p<0.05, * p<0.1.

Table F4: Baseline balance on households' characteristics

	HHs phone survey		
	N	Control mean	Treatment
	(1)	(2)	(3)
Households interviewed per CHW	227	3.47	0.18
			(0.18)
Household size	806	13.9	-0.31
			(0.68)
Pregnant woman (=1)	806	0.53	-0.046
			(0.037)
Child under 5 (=1)	805	0.95	-0.011
			(0.019)
Distance from CHW's house (minutes)	809	11.9	1.62
			(1.43)
Respondent's age	795	29.3	-0.39
			(0.55)
Household head (=1)	807	0.11	-0.020
			(0.022)
Muslim (=1)	805	0.43	-0.11*
			(0.063)
Education - 6 years or more	805	0.21	0.076**
			(0.034)
Married (=1)	805	0.77	-0.026
			(0.033)

Note. Column (1) reports the number of observations. Column (2) reports the sample mean of the control group. Column (3) reports estimates for the treatment indicator variable in equation (1). Standard errors, reported in parentheses, are clustered at the CHW level. *** p<0.01, ** p<0.05, * p<0.1.

G Additional measures on beliefs updating

Table G1: Difference between beliefs and health indicators

	CHWs face-to-face beliefs elicitation						
		Included in the intervention			Not included in the intervention		
	Average	U5 mortality	4 prenatal checks	Assisted birth	Birth at facility	Postnatal checks	Complete vaccines
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Treat X Post	-0.353***	0.080	-0.432***	-0.394***	-0.569***	-0.484***	-0.311***
	(0.078)	(.114)	(.139)	(.144)	(.135)	(.129)	(.115)
	[.001]	[.075]	[.003]	[.004]	[.001]	[.001]	[.004]
Control Mean	.001	.000	.000	.000	.000	.000	.000
Observations	484	481	484	484	484	484	484
Effect (%)	-32.43	19.40	-245.31	-22.32	-25.79	-33.59	-21.24
Control ρ	0.83***	0.73***	0.71***	0.67***	0.65***	0.64***	0.77***

Note: In this table, I report the intent-to-treat estimates of the direct effects of the treatment on service delivery to households. These are obtained by ordinary least squares (OLS) estimation of Equation 2. The estimation sample consists of CHWs who participated in the pre and post-intervention face-to-face beliefs elicitation. In all columns, the dependent variable is a z-score representing the value and sign of the difference between the elicited belief and the health indicator's correct value. The correct value is taken from MICS 2019. In column (1), the dependent variable is an index built using the procedure in Kling, Liebman, and Katz (2007). I calculate within-sample z-scores for each individual outcome, employing the mean and the standard deviation of the pure control group. I then obtain the unweighted average z-score for each group of outcomes. In columns (2)-(7), depending on the column, the relevant health indicator is described in the following. (2) 'In 2019, out of 100 children, how many would not survive to age 5?' (reversed). (3) 'In 2019, out of 100 pregnant women, how many completed the four recommended prenatal checks before delivery?'. (4) 'In 2019, out of 100 pregnant women, how many were giving birth attended by qualified health care personnel (doctors, nurses, or trained midwives)?'. (5) 'In 2019, out of 100 pregnant women, how many were giving birth at the health center?'. (6) 'In 2019, out of 100 new mothers, how many completed all recommended postnatal checks?'. (7) 'In 2019, out of 100 newborns, how many were receiving all the recommended vaccines on time?'. All specifications include CHW and household-level strata fixed effects. They also include CHW and household-level controls. The full list of controls is presented in Section 4. For each outcome, I report the mean outcome for the control group, the number of observations, the treatment effect in percentage points, and the autocorrelation (ρ) between pre and post-intervention outcome values. Standard errors are reported in parentheses and q-values in brackets, obtained using the sharpened procedure of Benjamini, Krieger, and Yekutieli (2006) a the table level. Standard errors are clustered at the CHW level. *** p<0.01, ** p<0.05, * p<0.1.

Table G2: Perceived program effectiveness

	CHWs face-to-face beliefs elicitation						
		Included in the intervention			Not included in the intervention		
	Average	U5 mortality	4 prenatal checks	Assisted birth	Birth at facility	Postnatal checks	Complete vaccines
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Treat X Post	-0.155**	0.007	-0.221	-0.217	-0.076	-0.302**	-0.116
	(0.077)	(0.136)	(0.158)	(0.146)	(0.150)	(0.148)	(0.127)
	[.187]	[.692]	[.274]	[.274]	[.431]	[.187]	[.431]
Control Mean	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Observations	484	483	484	484	484	484	484
R^2	0.217	0.204	0.155	0.175	0.196	0.166	0.214
Effect (%)	-2.59	-0.15	-3.33	-3.85	-1.41	-4.97	-1.97
Control ρ	0.77***	0.66***	0.43***	0.44***	0.30***	0.52***	0.59***

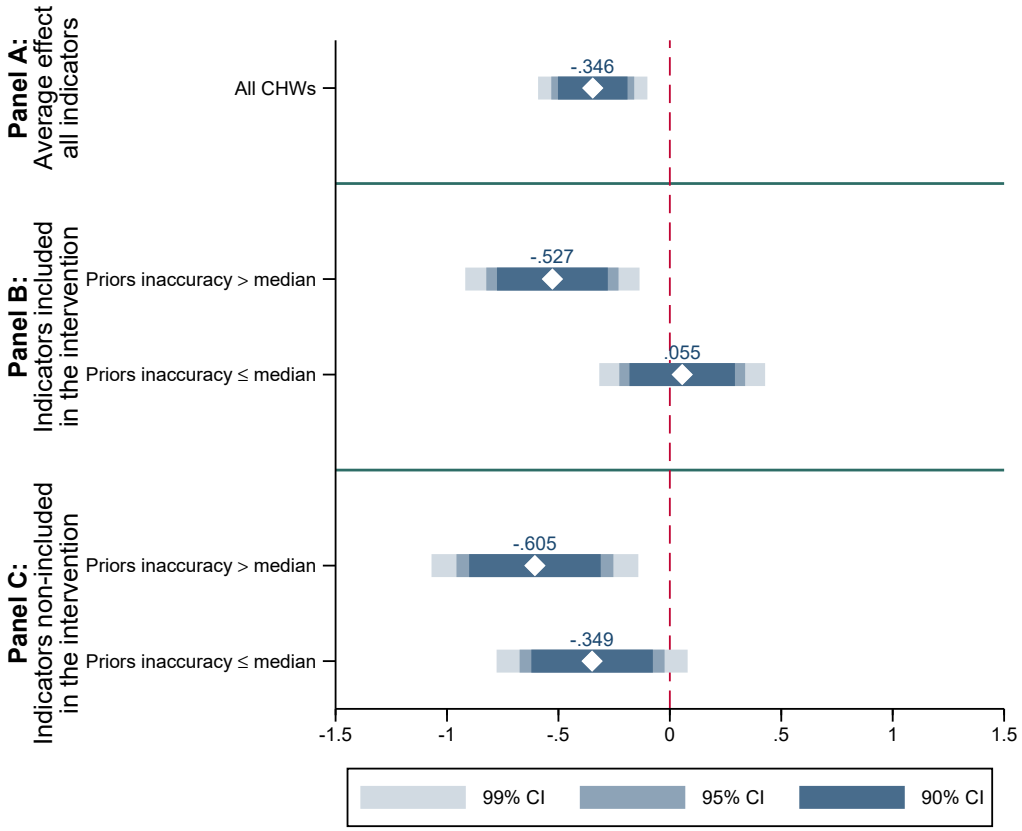
Note: In this table, I report the intent-to-treat estimates of the direct effects of the treatment on service delivery to households. These are obtained by ordinary least squares (OLS) estimation of Equation 2. The estimation sample consists of CHWs who participated in the pre and post-intervention face-to-face beliefs elicitation. In all columns, the dependent variable is a z-score representing the perceived program's effectiveness at improving health indicators at the local level. In column (1), the dependent variable is an index built using the procedure in Kling, Liebman, and Katz (2007). I calculate within-sample z-scores for each individual outcome, employing the mean and the standard deviation of the pure control group. I then obtain the unweighted average z-score for each group of outcomes. In columns (2)-(7), depending on the column, the relevant question is described in the following. (2) 'Do you think the CHWs program is improving the under-5 mortality indicator in the Oio and Farim region?'. (3) 'Do you think the CHWs program is improving the indicator of women completing the recommended 4 antenatal checks before delivery in the Oio and Farim region?'. (4) 'Do you think the CHWs program is improving the indicator of pregnant women giving birth attended by skilled health personnel in the Oio and Farim region?'. (5) 'Do you think the CHWs program is improving the indicator of pregnant women giving birth at the health center in the Oio and Farim region?'. (6) 'Do you think the CHWs program is improving the indicator of mothers completing all postnatal check-ups after delivery in the Oio and Farim region? '. (7) 'Do you think the CHWs program is improving the indicator of children completing all recommended vaccinations on time in the Oio and Farim region?'. All specifications include CHW-level strata fixed effects. They also include CHW-level controls. The full list of controls is presented in Section 4. For each outcome, I report the mean outcome for the control group, the number of observations, the treatment effect in percentage points, and the autocorrelation (ρ) between pre and post-intervention outcome values.. Standard errors are reported in parentheses and q-values in brackets, obtained using the sharpened procedure of Benjamini, Krieger, and Yekutieli (2006) a the table level. Standard errors are clustered at the CHW level. *** p<0.01, ** p<0.05, * p<0.1.

Table G3: Perceived program effectiveness, 4months post-intervention

	CHWs face-to-face survey and CHWs phone survey						
	Average	U5 mortality	4 prenatal checks	Assisted birth	Birth at facility	Postnatal checks	Complete vaccines
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Treat X Post	-0.107*	-0.001	-0.144	-0.142	-0.061	-0.217*	-0.078
	(0.060)	(0.110)	(0.115)	(0.119)	(0.121)	(0.114)	(0.106)
	[.539]	[.909]	[.598]	[.598]	[.909]	[.539]	[.909]
Treat X Follow-up (4 months)	-0.130	-0.035	-0.051	-0.288*	-0.230	-0.172	-0.006
	(0.111)	(0.175)	(0.164)	(0.153)	(0.154)	(0.157)	(0.171)
	[.598]	[.909]	[.909]	[.539]	[.598]	[.598]	[.909]
Control Mean	3.940	3.931	3.910	3.904	3.970	3.853	4.069
Observations	723	722	723	723	723	723	723
R^2	0.601	0.475	0.505	0.515	0.487	0.538	0.506
Effect - Post (%)	-2.72	-0.03	-3.68	-3.64	-1.54	-5.63	-1.92
Effect - Follow-up (%)	-3.30	-0.89	-1.30	-7.38	-5.79	-4.46	-0.15
Control ρ (Post, Pre)	0.77***	0.66***	0.43***	0.44***	0.30***	0.52***	0.59***
Control ρ (Follow-up, Pre)	0.18*	0.01	0.01	0.18*	0.09	0.15	0.11

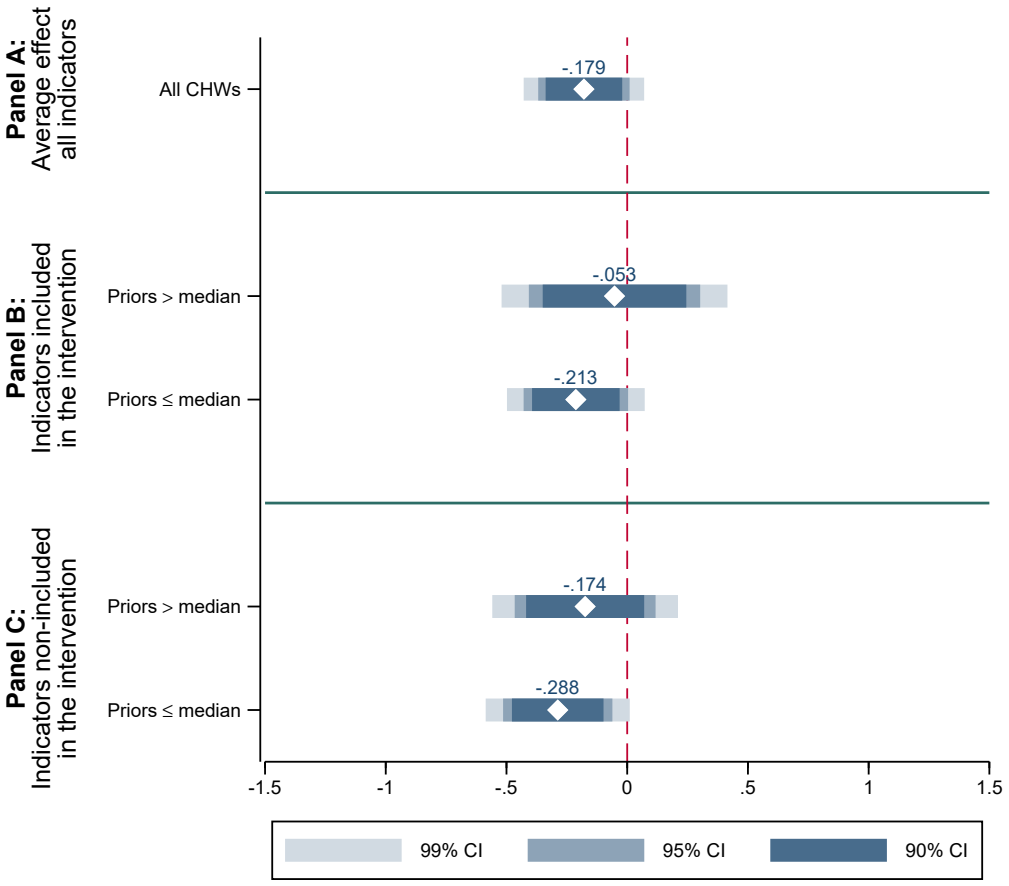
Note: In this table, I report the intent-to-treat estimates of the direct effects of the treatment on service delivery to households. These are obtained by ordinary least squares (OLS) estimation of Equation 2 for three periods, including an individual fixed effect at the level of the CHW. The estimation sample consists of all the CHWs interviewed in the follow-up phone survey. In all columns, the dependent variable is a z-score representing the perceived program's effectiveness at improving health indicators at the local level. In column (1), the dependent variable is an index built using the procedure in Kling, Liebman, and Katz (2007). I calculate within-sample z-scores for each individual outcome, employing the mean and the standard deviation of the pure control group. I then obtain the unweighted average z-score for each group of outcomes. In columns (2)-(7), depending on the column, the relevant question is described in the following. (2) 'Do you think the CHWs program is improving the under-5 mortality indicator in the Oio and Farim region?'. (3) 'Do you think the CHWs program is improving the indicator of women completing the recommended 4 antenatal checks before delivery in the Oio and Farim region?'. (4) 'Do you think the CHWs program is improving the indicator of pregnant women giving birth attended by skilled health personnel in the Oio and Farim region?'. (5) 'Do you think the CHWs program is improving the indicator of pregnant women giving birth at the health center in the Oio and Farim region?'. (6) 'Do you think the CHWs program is improving the indicator of mothers completing all postnatal check-ups after delivery in the Oio and Farim region? '. (7) 'Do you think the CHWs program is improving the indicator of children completing all recommended vaccinations on time in the Oio and Farim region?'. For each outcome, I report the mean outcome for the control group, the number of observations, the treatment effect in percentage points, and the outcome autocorrelation (ρ) between the periods indicated in parentheses. Standard errors are reported in parentheses and q-values in brackets, obtained using the sharpened procedure of Benjamini, Krieger, and Yekutieli (2006) a the table level. Standard errors are clustered at the CHW level. *** p<0.01, ** p<0.05, * p<0.1.

Figure G1: Change in distance between beliefs and health indicators - interaction with priors



Note. Estimates based on OLS regressions using Equation 3. The estimation sample consists of CHWs who participated in the pre and post-intervention face-to-face beliefs elicitation. I define 'inaccuracy' as the absolute value of the difference between the elicited belief and the health indicator's correct value. Panel A plots the effect of the treatment on change in the average belief inaccuracy for all the six indicators elicited. Panel B considers only the three indicators included in the intervention and plots the treatment effect on the change in the average belief inaccuracy for workers with prior beliefs on health indicators more or less inaccurate than the median. Panel C considers only the three health indicators not included in the intervention and plots the effect of the treatment on the change in the average belief inaccuracy for workers with prior beliefs more or less inaccurate than the median. All specifications include CHW-level strata fixed effects. They also include CHW-level controls. The full list of controls is presented in Section 4. Standard errors are clustered at the CHW level.

Figure G2: Change in perceived program effectiveness, 4months post-intervention - interaction with priors



Note. Estimates based on OLS regressions using Equation 3. The estimation sample consists of CHWs interviewed in the follow-up phone survey. Panel A plots the treatment effect on the change in the average perceived program's effectiveness in improving all the six indicators elicited. Panel B considers only the three indicators included in the intervention and plots the treatment effect on the change in the average perceived program's effectiveness for workers with prior perceptions about the program's effectiveness higher or lower than the median. Panel C considers only the three health indicators not included in the intervention and plots the treatment effect on the change in the average perceived program's effectiveness for workers with prior perceptions about the program's effectiveness higher or lower than the median. All specifications include CHW-level strata fixed effects. They also include CHW-level controls. The full list of controls is presented in Section 4. Standard errors are clustered at the CHW level.

H Households targeting

The increased healthcare time delivered to households may be compensated by CHWs targeting households living closer to their house or not fitting the program's target population. I find no evidence that targeting by distance or the presence of a pregnant woman or a child under 5 changes with treatment when looking at the visits during the last month (Table H1) or the past six months (Table H2). Turning to the behavior of the entire sample, the distance or the presence of a pregnant woman or a child under 5 changes are not correlated with the likelihood of receiving at least one visit during the last month (Table H3) or the past six months (Table H4). When looking at the number of visits for those households that received at least one visit, I find that during the last month, CHWs visited households with a child under 5 more often. Looking at the number of visits over the past six months, I see that CHWs visited more households with a child under five but less often further away households or households with a pregnant woman.

Table H1: Effect of treatment on household targeting - last month

	HHs phone survey		
	Pregnant woman in the household	Child under 5 in the household	Distance from CHW's house (minutes)
	(1)	(2)	(3)
Treat	-0.065	0.004	0.222
	(0.046)	(0.022)	(1.850)
Control Mean	0.535	0.937	11.717
Observations	623	622	622
R^2	0.105	0.098	0.090

Note: In this table, I report the intent-to-treat estimates of the direct effects of the treatment on service delivery to households. These are obtained by ordinary least squares (OLS) estimation of Equation 1. The estimation sample consists of women of fertile age (18-49 years old) interviewed in the follow-up phone survey who received at least one visit the month before the interview. Depending on the column, the dependent variables are defined by the following. (1) Indicator variable that takes value of 1 if respondent answered 'Yes' to the question: 'Are there any pregnant women in your household?', and 0 otherwise. (2) Indicator variable that takes value of 1 if respondent answered 'Yes' to the question: 'Are there any children under the age of 5 in your household?', and 0 otherwise. (3) Continuous variable that indicates how long it would take the respondent in minutes to walk to the house of the CHW identified by the neighborhood and house directions provided by the CHW. All specifications include CHW-level strata fixed effects. They also include CHW and household-level controls. The full list of controls is presented in Section 4. Standard errors are reported in parentheses. Standard errors are clustered at the CHW level. *** p<0.01, ** p<0.05, * p<0.1.

Table H2: Effect of treatment on household targeting - past six months

	HHs phone survey		
	Pregnant woman in the household	Child under 5 in the household	Distance from CHW's house (minutes)
	(1)	(2)	(3)
Treat	-0.027	0.000	0.218
	(0.040)	(0.018)	(1.659)
Control Mean	0.526	0.945	11.833
Observations	791	790	790
R^2	0.088	0.100	0.070

Note: In this table, I report the intent-to-treat estimates of the direct effects of the treatment on service delivery to households. These are obtained by ordinary least squares (OLS) estimation of Equation 1. The estimation sample consists of women of fertile age (18-49 years old) interviewed in the follow-up phone survey that received at least one visit in the six months previous to the interview. Depending on the column, the dependent variables are defined by the following. (1) Indicator variable that takes value of 1 if respondent answered 'Yes' to the question: 'Are there any pregnant women in your household?', and 0 otherwise. (2) Indicator variable that takes value of 1 if respondent answered 'Yes' to the question: 'Are there any children under the age of 5 in your household?', and 0 otherwise. (3) Continuous variable that indicates how long it would take the respondent in minutes to walk to the house of the CHW identified by the neighborhood and house directions provided by the CHW. All specifications include CHW-level strata fixed effects. They also include CHW and household-level controls. The full list of controls is presented in Section 4. Standard errors are reported in parentheses. Standard errors are clustered at the CHW level. *** p<0.01, ** p<0.05, * p<0.1.

Table H3: Household characteristics and receiving at least one visit – last month

	HHs phone survey		
	Pregnant woman in the household	Child under 5 in the household	Distance from CHW's house (minutes)
	(1)	(2)	(3)
At least one visit	-0.007	0.003	-0.397
	(0.047)	(0.023)	(1.540)
Mean	0.504	0.939	12.759
Observations	806	805	809
R^2	0.071	0.077	0.066

Note: In this table, I report the correlation between household characteristics and whether it received at least one visit during the last month. These are obtained by ordinary least squares (OLS) estimation of Equation 1. The estimation sample consists of women of fertile age (18-49 years old) interviewed in the follow-up phone survey. Depending on the column, the dependent variables are defined by the following. (1) Indicator variable that takes value of 1 if respondent answered 'Yes' to the question: 'Are there any pregnant women in your household?', and 0 otherwise. (2) Indicator variable that takes value of 1 if respondent answered 'Yes' to the question: 'Are there any children under the age of 5 in your household?', and 0 otherwise. (3) Continuous variable that indicates how long it would take the respondent in minutes to walk to the house of the CHW identified by the neighborhood and house directions provided by the CHW. All specifications include CHW and household-level strata fixed effects as presented in Section 4. Standard errors are reported in parentheses. Standard errors are clustered at the CHW level. *** p<0.01, ** p<0.05, * p<0.1.

Table H4: Household characteristics and receiving at least one visit – past six months

	HHs phone survey		
	Pregnant woman in the household	Child under 5 in the household	Distance from CHW's house (minutes)
	(1)	(2)	(3)
At least one visit	0.182	0.081	-8.805
	(0.197)	(0.115)	(6.861)
Mean	0.504	0.939	12.759
Observations	806	805	809
R^2	0.073	0.078	0.068

Note: In this table, I report the correlation between household characteristics and whether it received at least one visit during the last six months. These are obtained by ordinary least squares (OLS) estimation of Equation 1. The estimation sample consists of women of fertile age (18-49 years old) interviewed in the follow-up phone survey. Depending on the column, the dependent variables are defined by the following. (1) Indicator variable that takes value of 1 if respondent answered 'Yes' to the question: 'Are there any pregnant women in your household?', and 0 otherwise. (2) Indicator variable that takes value of 1 if respondent answered 'Yes' to the question: 'Are there any children under the age of 5 in your household?', and 0 otherwise. (3) Continuous variable that indicates how long it would take the respondent in minutes to walk to the house of the CHW identified by the neighborhood and house directions provided by the CHW. All specifications include CHW and household-level strata fixed effects as presented in Section 4. Standard errors are reported in parentheses. Standard errors are clustered at the CHW level. *** p<0.01, ** p<0.05, * p<0.1.

I Additional measures of health knowledge

This section reports additional measures of treatment effects on learning about health practices and knowledge diffusion. Table I1 shows estimates of treatment effects on CHWs' learning of health practices over time by considering the reduced version of the knowledge test. I find an overall positive treatment effect consistent with Hypothesis 2a. In particular, the first row reports the grade change registered between the pre and post-intervention knowledge tests during the workshop, considering only the ten questions comprising the reduced test. CHWs in the treatment group show an overall ten percent larger improvement between pre and post-intervention tests, significant at the 1 percent level, compared to the control group (column 1). The different test components show that this effect is concentrated in the module assessing CHWs' knowledge of 'Danger signs in pregnant women' (column 4). The magnitude of the effect is 18 percent, significant at the 1 percent level. I do not find significant treatment effects on the other components of the test, namely 'Antenatal care' (column 2) and 'HIV prevention and testing' (column 3). The second row reports the grade change registered between the pre-training test and follow-up phone interview four months after the intervention. I do not see any significant treatment effect, although the sign and magnitude of the coefficient are qualitatively consistent with the first row. As for the complete version of the test, it is important to note the unbalance in the pre-training test grades as reported in Table I3. Finally, I do not find any evidence of downstream effects regarding knowledge diffusion at the household level, as reported in Table I2.

Table I1: Change in grades from pre to post-test and phone follow-up - reduced test

	CHWs written tests and CHWs phone survey			
	Full test	Antenatal care	HIV prevention and testing	Danger signs in pregnant women
	(1)	(2)	(3)	(4)
Treat X Post-test	0.082***	0.049	0.002	0.154***
	(0.037)	(0.038)	(0.052)	(0.063)
	[.122]	[.244]	[.354]	[.122]
Treat X Follow-up (4 months)	0.055	0.068	-0.022	0.080
	(0.040)	(0.048)	(0.055)	(0.061)
	[.244]	[.244]	[.354]	[.244]
Control Mean	0.843	0.814	0.874	0.858
Observations	727	727	727	727
R^2	0.579	0.558	0.454	0.539
Effect - Post-test (%)	9.73	6.02	0.23	17.95
Effect - Follow-up (%)	6.52	8.35	-2.52	9.32
Control ρ (Post, Pre)	0.68***	0.62***	0.23**	0.44***
Control ρ (Follow-up, Pre)	0.11*	0.01	-0.05	0.05

Note: In this table, I report the intent-to-treat estimates of the direct effects of the treatment on learning. These are obtained by ordinary least squares (OLS) estimation of Equation 2 for three periods, including an individual fixed effect at the level of the CHW. The estimation sample consists of all the CHWs interviewed in the follow-up phone survey. Depending on the column, the dependent variables are defined by the following. (1) Continuous variable indicates the grade over the entire reduced version of the test, containing ten questions, and takes values from 0 to 1. (2) Continuous variable that indicates the grade over four questions focused on antenatal checks during pregnancy and takes values from 0 to 1. (3) Continuous variable that indicates the grade over two questions focused on HIV testing during pregnancy and takes values from 0 to 1. (4) Continuous variable that indicates the grade over four questions focused on the danger signs during pregnancy and takes values from 0 to 1. For each outcome, I report the mean outcome for the control group, the number of observations, the treatment effect in percentage points, and the outcome autocorrelation (ρ) between the periods indicated in parentheses. Standard errors are reported in parentheses and q-values in brackets, obtained using the sharpened procedure of Benjamini, Krieger, and Yekutieli (2006) a the table level. Standard errors are clustered at the CHW level. *** p<0.01, ** p<0.05, * p<0.1.

Table I2: Households knowledge

	HHs phone survey			
	Full test	Antenatal care	HIV prevention and testing	Danger signs in pregnant women
	(1)	(2)	(3)	(4)
Treat	-0.006	-0.018	0.015	-0.004
	(0.009)	(0.014)	(0.021)	(0.007)
	[1]	[1]	[1]	[1]
Control Mean	0.861	0.805	0.758	0.968
Observations	800	799	792	800
R^2	0.065	0.083	0.059	0.069
Effect (%)	-0.70	-2.24	-1.98	-0.41

Note: In this table, I report the intent-to-treat estimates of the direct effects of the treatment on households' knowledge of health practices promoted by the program. These are obtained by ordinary least squares (OLS) estimation of Equation 1. The estimation sample consists of women (18-49 years old) interviewed in the follow-up phone survey. Depending on the column, the dependent variables are defined by the following. (1) Continuous variable indicates the grade over the entire reduced version of the test, containing ten questions, and takes values from 0 to 1. (2) Continuous variable that indicates the grade over four questions focused on antenatal checks during pregnancy and takes values from 0 to 1. (3) Continuous variable that indicates the grade over two questions focused on HIV testing during pregnancy and takes values from 0 to 1. (4) Continuous variable that indicates the grade over four questions focused on the danger signs during pregnancy and takes values from 0 to 1. All specifications include CHW and household-level strata fixed effects. They also include CHW and household-level controls. The full list of controls is presented in Section 4. Standard errors are reported in parentheses and q-values in brackets, obtained using the sharpened procedure of Benjamini, Krieger, and Yekutieli (2006) a the table level. Standard errors are clustered at the CHW level. *** p<0.01, ** p<0.05, * p<0.1.

Table I3: Pre-refresher training grades

	CHWs written test		
	N	Control Mean	Treatment
	(1)	(2)	(3)
Full test	245	0.63	-0.078***
			(0.022)
Antenatal care	245	0.63	-0.057**
			(0.024)
HIV prevention and testing	245	0.62	-0.073**
			(0.032)
Danger signs in pregnant women	245	0.64	-0.13***
			(0.037)

Note. Column (1) reports the number of observations. Column (2) reports the sample mean of the control group. Column (3) reports estimates for the treatment indicator variable in equation (1). Standard errors, reported in parentheses, are clustered at the CHW level. *** p<0.01, ** p<0.05, * p<0.1.

Table I4: Pre-refresher training grades – reduced version of the test

	CHWs written test		
	N	Control Mean	Treatment
	(1)	(2)	(3)
Full test	245	0.77	-0.11***
			(0.029)
Antenatal care	245	0.79	-0.10***
			(0.035)
HIV prevention and testing	245	0.81	-0.041
			(0.039)
Danger signs in pregnant women	245	0.73	-0.15***
			(0.044)

Note. Column (1) reports the number of observations. Column (2) reports the sample mean of the control group. Column (3) reports estimates for the treatment indicator variable in equation (1). Standard errors, reported in parentheses, are clustered at the village level. *** p<0.01, ** p<0.05, * p<0.1.

J Robustness Checks

Table J1: Heterogeneous treatment effects on service delivery by household sampling strata

	HHs phone survey						
	Past month			Past six months			
	Visits received	Visit duration (minutes)	Time with the CHW (minutes)	Visits received	Visit duration (minutes)	Time with the CHW (minutes)	Satisfaction with the program
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Treat	-0.005	5.481***	11.631**	-0.071	4.721**	28.724	0.188*
	(0.106)	(1.778)	(4.867)	(0.375)	(1.890)	(17.959)	(0.100)
Treat X Nudged	0.223	-4.212	-3.981	-0.188	-2.243	-25.446	-0.117
	(0.152)	(2.690)	(7.366)	(0.599)	(2.871)	(28.175)	(0.162)
Treat X Contacted	0.475**	-4.800	-0.502	0.903	-3.995	23.195	-0.024
	(0.217)	(3.093)	(8.478)	(0.676)	(4.217)	(38.899)	(0.210)
Control Mean	1.317	15.069	26.522	5.270	23.516	132.713	0.000
Observations	786	782	782	789	786	786	795
R^2	0.162	0.133	0.126	0.139	0.157	0.207	0.075

Note. In this table, I report the intent-to-treat estimates of the direct effects of the treatment on service delivery to households. These are obtained by ordinary least squares (OLS) estimation of Equation 1. The estimation sample consists of women (18-49 years old) interviewed in the follow-up phone survey. Depending on the column, the dependent variables are defined by the following. (1) Number of visits received during November 2022. (2) Average duration of a visit during November 2022. (3) The number of visits during November 2022 multiplied by the average duration of a visit during November 2022. (4) Visits received since June 2022. (5) Average duration of a visit since June 2022. (6) The number of visits since June 2022 multiplied by the average duration of a visit since June 2022. (7) A 5-item Likert scale that takes the values 1 'Not at all,' 2 'A little,' 3 'Moderately,' 4 'A lot,' and 5 'Extremely' to measure replies to the following question: 'How satisfied are you with the CHWs program?' The first row reports the treatment effect coefficients for the stratum 'No condition.' The second row reports the coefficients of the treatment dummy interacted with the stratum 'Flagged at CHWs follow-up.' The third row reports the coefficients of the treatment dummy interacted with the stratum 'Contacted during the workshop.' The strata are illustrated in Figure D2. All specifications include CHW strata fixed effects. They also include CHW and household-level controls. The full list of controls is presented in Section 4. Standard errors are clustered at the CHW level. *** p<0.01, ** p<0.05, * p<0.1.

K Cost effectiveness of the intervention

Table K1: Costs-benefit analysis

	US Dollars 2022	US Dollars 2022 PPP
	(1)	(2)
Total cost	361.72	1,035.83
Cost per worker	2.91	8.35
Cost per additional SD	5.56	15.96

Note: Column (1) reports the values in US Dollars in 2022 when the intervention was implemented (World Development Indicators, The World Bank, Indicator Code: PA.NUS.FCRF). Column (2) corrects for Purchase Power Parity (World Development Indicators, The World Bank, Indicator Code: PA.NUS.PPP). To compute these values, I followed the methodology described in (Dhaliwal et al., 2013).