UC Berkeley
CEGA Working Papers

Title

Building Resilient Health Systems: Experimental Evidence from Sierra Leone and the 2014 Ebola Outbreak

Permalink

Authors

Christensen, Darin
Dube, Oeindrila
Haushofer, Johannes
et al.

Publication Date

DOI

Building Resilient Health Systems: Experimental Evidence from Sierra Leone and the 2014 Ebola Outbreak

Darin Christensen, Oeindrila Dube, Johannes Haushofer, Bilal Siddiqi and Maarten Voors

Working Paper Series

Center for Effective Global Action
University of California

eScholarship
University of California

The CEGA Working Paper Series showcases ongoing and completed research by faculty affiliates of the Center. CEGA Working Papers employ rigorous evaluation techniques to measure the impact of large-scale social and eco-nomic development programs, and are intended to encourage discussion and feedback from the global development community.

BUILDING RESILIENT HEALTH SYSTEMS:
EXPERIMENTAL EVIDENCE FROM SIERRA LEONE AND THE 2014 EBOLA OUTBREAK[*]

Darin Christensen[†]
Oeindrila Dube[‡]
Johannes Haushofer[§]
Bilal Siddiqi[¶]
Maarten Voors[‖]

March 26, 2020

Abstract

Developing countries are characterized by high rates of mortality and morbidity. A potential contributing factor is the low utilization of health systems, stemming from the low perceived quality of care delivered by health personnel. This factor may be especially critical during crises, when individuals choose whether to cooperate with response efforts and frontline health personnel. We experimentally examine efforts aimed at improving health worker performance in the context of the 2014–15 West African Ebola crisis. Roughly two years before the outbreak in Sierra Leone, we randomly assigned two accountability interventions to government-run health clinics — one focused on community monitoring and the other gave status awards to clinic staff. We find that over the medium run, prior to the Ebola crisis, both interventions led to improvements in utilization of clinics and patient satisfaction with the health system. In addition, child health outcomes improved substantially in the catchment areas of community monitoring clinics. During the crisis, the interventions also led to higher reported Ebola cases, as well as lower mortality from Ebola — particularly in areas with community monitoring clinics. We explore three potential mechanisms: the interventions (1) increased the likelihood that patients reported Ebola symptoms and sought care; (2) unintentionally increased Ebola incidence; or (3) improved surveillance efforts. We find evidence consistent with the first: by building trust and confidence in health workers, and improving the perceived quality of care provided by clinics prior to the outbreak, the interventions encouraged patients to report and receive treatment. Our results suggest that accountability interventions not only have the power to improve health systems during normal times, but can additionally make health systems resilient to crises that may emerge over the longer run.

1. Introduction

Over 8 million people die annually in low- and middle-income countries from treatable conditions. In addition to the human suffering, in 2015 these preventable deaths generated USD 6 trillion in economic losses (Kruk et al. 2018). These deaths are particularly tragic because care is often available at little to no cost, but there is under-utilization of available health services (Dupas 2011). Low quality of care is thought to be a major factor responsible for low utilization (Banerjee et al. 2004; Das et al. 2016). More than half of patients surveyed by the Lancet Global Health Commission across 12 countries report that they did not seek necessary medical care in the preceding year, because they lacked confidence in their local health system (Kruk et al. 2018, e1211).

This point has been underscored by the recent intractable Ebola outbreaks — in West Africa from 2014–2016, and in the Democratic Republic of Congo since 2018. In September 2014, the World Health Organization (WHO) described West Africa's Ebola epidemic as "the most severe acute public health emergency seen in modern times. Never before in recorded history has a biosafety level four pathogen infected so many people, so quickly, over such a broad geographic area, for so long" (WHO 2014). At that point, less than 7,000 individuals had been infected. By the end of the crisis in early 2016, the Centers for Disease Control and Prevention (CDC) estimated more than 28,000 confirmed, suspected, or probable cases (CDC 2019). Sierra Leone, one of the three most heavily impacted countries, accounts for roughly half of those cases and just under 4,000 deaths.

As with other infectious diseases, Ebola containment efforts emphasize early isolation and treatment. Ebola spreads through the transmission of infected bodily fluids; reducing the reproduction rate, thus, requires "reducing the time between when people first show symptoms and are isolated" (Whitty et al. 2014, 193).[1] Yet, fears about sub-standard care and a lack of trust in health workers and the health system deterred symptomatic patients from visiting clinics: "Local communities were suspicious of efforts to test, treat, and isolate patients with Ebola symptoms and engaged in practices of hiding sick family members, running away from local communities, or attempting to manage the course of Ebola within local households and communities" (Abramowitz et al. 2016, 24). Thus, low utilization of health services, reflecting a lack of confidence in the quality of care, is thought to have been a major obstacle to early treatment and containment of the Ebola epidemic in West Africa. This is not specific to Ebola: public confidence in health systems affects efforts to contain other epidemics, including the novel coronavirus (COVID-19).

What can be done to improve confidence in health workers, and the perceived quality of care of health

---

[1] Seeking treatment not only prevents transmission, but also increases the patient's survival prospects: Garske et al. (2017, 5) report case fatality rates of over 90 percent for patients in Sierra Leone with no or unknown hospitalization status, but that rate drops to less than 60 percent for patients admitted to Ebola treatment units or holding centers in Sierra Leone (see also Waxman et al. 2017).

systems, to make them more resilient to crises? The Lancet Global Health Commission focuses attention on two groups: patients — who must be "accountability agents, able to hold health system actors to account" — and providers: "Demotivated providers cannot contribute to a high-quality health system...[Resilience] requires accountable leaders who respect and motivate their frontline staff" (Kruk et al. 2018, e1200).

These recommendations echo claims in personnel economics about how to motivate difficult-to-monitor frontline workers (Finan et al. 2017). While the principal-agent model makes clear how financial incentives can induce effort, performance pay may not always be feasible under resource constraints. Moreover, if frontline health workers perceive intrinsic benefits from their work, financial rewards may crowd out intrinsic motivation (Benabou and Tirole 2003; Besley and Ghatak 2005; Dixit et al. 2002). Fortunately, non-financial approaches to motivating workers also exist. Organizations can harness social incentives that arise from interactions between providers and clients, or among providers themselves (Besley and Ghatak 2005; Ashraf and Bandiera 2018). For example, competition can be engendered among health workers as an alternate route for improving effort. Empowering citizens with monitoring tools to hold providers accountable can serve as another lever (Mansuri and Rao 2003).

We evaluate these two ideas using a field experiment implemented just before the Ebola outbreak in Sierra Leone.[2] In partnership with the Government of Sierra Leone (GoSL) and 3 international NGOs, we randomly assigned two interventions (and control) to 254 government-run health clinics.[3] The timing of the experiment relative to the Ebola epidemic affords us a unique opportunity to measure the effects of these interventions under business-as-usual and crisis conditions. Unlike past studies, we can observe whether the interventions contribute to the health system's resilience — a capacity to respond to crises and changing population needs that we observe only when a system faces an adverse shock.

The first intervention promotes social accountability, providing patients with the information and forum to confront frontline health providers. Modeled on a program evaluated by Björkman and Svensson (2009), the community monitoring (CM) intervention creates scorecards ranking local health services; convenes interface meetings with community members and clinic staff to discuss these ratings and to develop "joint action plans" to improve service delivery; and follows up with meetings to monitor progress after 1, 3, and 9 months. The second intervention motivates clinic staff through a competition that provides non-financial awards (NFA) to the best and most improved clinic in each district. The competition is advertised to clinic staff, who are also revisited three times during implementation to sustain interest. The winners receive a letter of commendation from a high-ranking politician and a plaque or wall clock for the clinic. The interventions do not provide physical inputs to clinics; rather, they attempt to empower patients to demand, and motivate nurses to supply, higher quality care despite resource constraints.

---

[2] Endline surveying for our evaluation concluded in June 2013; the first Ebola case was reported in May 2014 (see Figure 2).

[3] The interventions were funded by the World Bank and implemented by Concern Worldwide, the International Rescue Committee, and Plan International with support from GoSL's Decentralization Secretariat and the Ministry of Health and Sanitation.

We leverage this experimental design and original data from 5,080 households, 508 community leaders, and 254 clinics to estimate the causal effects of these interventions on both medium-run outcomes measured one year later, as well as their longer run effects on reporting during the Ebola crisis. The clinics in our study cover nearly 1 million people, just over 15 percent of Sierra Leone's population in 2011.

In the medium-run, prior to the crisis, we find that both interventions improve the perceived quality of care and confidence in the health system, as reflected in individuals' satisfaction with health workers and their increased utilization of clinics. Facing a health need, individuals living in treatment areas are 7 percent more likely to seek care in a government-run clinic; and satisfaction increases by 0.1 standard deviations. These effects are similar across both the CM and NFA arms. In CM alone, we find increases in maternal utilization — for example, the probability of delivery in a facility increases by 11 percent — as well as improvements in child health outcomes. The likelihood of under-5 death in the household declines by 38 percent, which we attribute to maternal utilization, as well as increases in vaccinations and improvements in child weight-for length. These effects are similar in magnitude to Björkman and Svensson (2009), who find a 33 percent reduction in under-5 mortality in clinics subject to community monitoring in Uganda. In our experimental context, neither intervention changes the quantity of services provided, nor leads to greater resources at the clinic level. This is not surprising, as the interventions do not attempt to increase inputs to clinics from higher levels of government.

To examine longer-run effects on outcomes related to the Ebola crisis, we use a de-identified database maintained by the GoSL and CDC. We count the number of reported cases (including individuals who test negative for Ebola) in a given week and section — a small administrative unit with a median area of 40 square kilometers.[4] Pooling the treatments, we estimate that the interventions increased total reported cases substantially, by just over 60 percent. While both treatments generate sizable increases, and we cannot reject the null that the treatments have equivalent effects, we see a larger increase in total reported cases in CM (above 70 percent). While most cases test negative for Ebola, the interventions also increase the number of infected patients reporting: a back-of-the-envelope calculation suggests that the effects on reporting by infected individuals reduces the disease's reproduction rate (R_0) by around 19 percent (Pronyk et al. 2016). We find no evidence that geographic spillovers, i.e. the movement of patients from control to treatment sections, amplify these effects.

Analogous to our medium-run effects, improvements in health outcomes during the Ebola crisis concentrate in areas receiving the CM intervention. Among patients who report, we find that a smaller number die in treated sections: our results imply that 1 patient dies for every 7 that reported in treated sections in the current or last week; that worsens to 1 in every 4 in control. This difference is driven by sections with CM

---

[4] The 254 clinics in the full sample fall in 205 sections. 160 sections contain a single clinic and, thus, have a unique treatment status and common dosage. We use this as our primary sample in analyzing Ebola-related outcomes; yet, we find similar effects with a dose-response model that employs the full sample of 205 sections (see Table E.11).

clinics.

We attribute the effects on Ebola reporting to improved perceptions in the quality of care and greater trust and confidence in the government health system. The medium-run effects on satisfaction and utilization are also present in the subset of clinics used in our Ebola analysis. Moreover, we find that individuals in both treatments express greater confidence in western medicine in our endline survey relative to traditional and spiritual healers, the principal alternatives to government clinics in rural Sierra Leone. We combine our measures of utilization, satisfaction with public health workers, and the relative efficacy of western medicine, into a perceived quality of care index. Instrumenting that index with our treatment, we find that a one standard deviation change in the perceived quality of care increases total reported cases by 0.43 per section-week.

We do not find support for two alternative explanations. First, the interventions do not appear to have increased actual Ebola transmission. We find that all types of cases, including patients with both confirmed and negative test results, increase in sections with treated clinics. As a consequence, the ratio of confirmed to total cases does not change with treatment. To rule out contagion within clinics, we investigate whether confirmed cases test positive due to exposure that may have occurred upon reporting (e.g., through contact in a waiting room). We can rule out such nosocomial transmission (i.e., exposure originating in a clinical setting) in 99 percent of cases, based on the timing symptom onset and reporting. Second, we also do not find that sections with treated clinics have more resources for top-down surveillance: sections with treated clinics do not host more facilities specializing in Ebola care, and there are no differences in laboratory testing or case workers. If anything, we find that a higher proportion of confirmed cases undergo contact tracing (i.e., the process of identifying recent contacts to flag at-risk individuals) in sections with control clinics.

Our results highlight two important points. First, accountability interventions can improve health systems and health outcomes by increasing the perceived quality of care, and by building trust and confidence in health providers. Second, perceived quality of care and confidence in the health system may become especially important during crises, when individuals face a choice about whether to cooperate with response efforts, such as heeding an evacuation order, honoring a quarantine, or voluntarily reporting for medical testing. Experts argue that low confidence in health systems undermines efforts to contain the current outbreak of the novel coronavirus (COVID-19), which has already infected over 95,000 individuals globally.[5] In addition, trust in government has been shown to be correlated with the decision to utilize care during the Ebola crises of Liberia (Blair et al. 2017; Morse et al. 2016; Tsai et al. 2019) and the DRC (Vinck et al. 2019) — claims which mirror recent evidence showing that fear and distrust deter patients from utilizing

---

[5] Wen (2020) writes, "A robust response [to COVID-19] from medical and public health practitioners has already begun. But for any response to be effective, people need to heed government officials' orders, and for that, they must have faith that their leaders know what they're doing and have the citizens' best interests at heart." In China, fears of under-reporting have led officials to adopt intrusive monitoring strategies (Mozur 2020) and begin offering large cash payments to patients who report, with one city reportedly offering over 1,400 USD to a patient who later tests positive (Bostock 2020).

4

health systems (Alsan and Wanamaker 2017; Blair et al. 2017; Vinck et al. 2019; Lowes and Montero 2018). As such, accountability interventions may confer particularly large benefits during crisis conditions, through their effects on reporting and cooperation, which can help reduce mortality and the spread of an epidemic.

To the best of our knowledge, we present the first experimental evidence showing this potential benefit, by demonstrating how accountability interventions can improve improve perceptions of the health system, increase epidemic reporting, and lower mortality conditional on reporting during a crisis. In this regard, our findings shed light on how accountability interventions can make health systems more resilient to major disruptions such as a health crisis.[6] Other evaluations of social accountability programs may miss this potential benefit given the infrequency of crises events.

More broadly, our findings bear on a larger literature on how to increase accountability among frontline bureaucrats and improve public services. The effects we observe do not align with net crowd-out of intrinsic motivation. Instead, they are consistent with previous work suggesting that social incentives such as non-financial awards can be used to boost worker performance in mission-oriented settings (Ashraf et al. 2014), as well as other experimental settings (Kosfeld and Neckermann 2011) and the private sector (Markham et al. 2002). Empowering citizens through community monitoring has also been examined as a tool for improving service delivery in a variety of social sectors, including education (Banerjee et al. 2010; Pradhan et al. 2011; Barr et al. 2012; Andrabi et al. 2018), corruption (Fiala and Premand 2018; Olken 2007) and health. As discussed above, our results align closely with Björkman and Svensson (2009), but are further from the more recent work of Raffler et al. (2019) who also examine community monitoring of health clinics in Uganda, but find weak effects. A possible reason for these contrasting results may lie in the baseline health conditions under which the interventions were implemented. The Björkman and Svensson (2009) intervention was implemented in 2004-2005, when under-5 mortality in Uganda stood at 117 per 1,000 live births. Analogously, our accountability interventions in Sierra Leone were implemented when under-5 mortality stood at 149.8 per 1,000 live births. In contrast, the more recent study in Uganda implemented community monitoring over 2014-2016, when infant mortality had fallen to 59 per 1,000 live births. The lack of consistent effects in the latter context suggest that it may be more difficult to improve health outcomes when baseline conditions themselves have already improved substantially (Raffler et al. 2019).[7]

The interventions we analyze harness social accountability and competition between providers to improve interactions between providers and the communities they serve. Other related approaches to improving health service delivery have examined the effects of financial incentives (Miller et al. 2012; Olken et al. 2014; Singh and Mitra 2017); combined technological monitoring with financial incentives (Banerjee et al.

---

[6] Relatedly, Bandiera et al. (2019) find that an empowerment program for young women in Sierra Leone increases their capacity to cope with disruptions caused by the Ebola crisis.

[7] When implemented under challenging baseline conditions, the effects of community monitoring appear to persist. A follow-up study of the original Björkman and Svensson (2009) intervention suggests a lack of convergence in health outcomes up to four years later (Björkman Nyqvist et al. 2017).

2008); and leveraged social incentives among patients, for example, through social signaling (Karing 2019). Our paper is also related to a larger literature that examines various approaches to improving health outcomes in the developing world. A complete review of this literature can be found in Dupas (2011) and Dupas and Miguel (2017).

The rest of our paper is structured as follows. Section 2 describes the experimental design and details of the two interventions. Section 3 introduces our sampling procedure, the survey and Ebola case data, randomization and empirical strategy. Section 4 presents our medium-run findings and plausible mechanisms, longer run Ebola outcomes, and investigates alternative explanations. The final section concludes with a discussion of policy implications.

2. Healthcare in Sierra Leone

2.1 Background

According to the World Development Indicators, Sierra Leone had the second highest under-5 mortality rate in the world in 2011, with 149.8 deaths per 1,000 live births. In the same year, the country had the highest maternal mortality rate in the world. Figure A.2 displays Sierra Leone's per capita health expenditure and under-5 mortality in 2010 relative to other countries that the World Bank classified as low income. Located in the upper-right quadrant, the country spent more and performed worse than countries at a comparable level of economic development.

In an effort to reduce infant and maternal mortality, the Government of Sierra Leone (GoSL) launched a free health care initiative in 2010, removing fees for pregnant women and children under 5 years old. The policy simultaneously increased pay for government healthcare workers; at the time 30–50 percent of staff did not receive a government wage and instead relied on charging illegal fees or inflated drug prices and accepting in-kind contributions from the communities they served. In addition to removing cost barriers and severe resource constraints, the GoSL saw a need to strengthen incentives for front-line healthcare workers. Absent incentives tied to service delivery, the government worried that nurses would miss work or continue to charge illegal fees — barriers to care that the free healthcare initiative was intended to eliminate.

Government-run clinics and hospitals are the primary providers of western-style medicine in Sierra Leone; private and NGO-sponsored facilities are scant (Denney and Mallett 2014). These facilities operate alongside traditional birth attendants and healers. Government clinics, referred to as peripheral health units (PHUs), come in three types: maternal and child health posts (MCHPs), community health posts (CHPs),

and community health centers (CHCs). The MCHPs and CHPs that comprise our study sample are primary health facilities — the first points of contact for patients in towns and villages — that each service populations of 500 to 10,000 (UNICEF 2014, 5). The 192 MCHPs and 62 CHPs included in our study covered a population close to 1 million people, nearly 15 percent of the country's population of 6.6 million in 2011.

PHUs continued to operate during the Ebola crisis: a UNICEF (2014, 17) facility survey in October 2014 (four months after the first confirmed case in Sierra Leone) found that only 4 percent of PHUs were closed. Levy et al. (2015, 753) report that "early assessments [from October 2014] found that many [Ebola] patients were first seeking care at local PHUs." Concerned that these PHUs lacked the training and equipment to properly isolate and care for Ebola patients, PHU staff were rapidly trained on infection prevention and control and outfitted with personal protective equipment. By early December 2014, 81 percent of health-care workers in Sierra Leone had received training (see Table E.2); by early January, training had reached 98 percent (Levy et al. 2015). In addition to screening and providing "no-touch" treatment for dehydration and fever, a case study in Kenema District found that PHU staff (specifically, community health workers) were also engaged in social mobilization, contact tracing, and community-based surveillance (Vandi et al. 2017). While training and the disbursement of protective equipment filled important knowledge and resource gaps, UNICEF's (2014, 17) survey found that 90 percent of PHUs felt that fear and misconceptions were "the main challenge confronted by the health system in fighting Ebola."

2.2 Interventions

2.2.1 Community Monitoring

We model our Community Monitoring (CM) intervention on Björkman and Svensson's (2009) "Power to the People" intervention, which was implemented in Uganda in 2005. The intervention attempts to mobilize "client power," empowering patients with the information and forum to demand accountability from frontline staff (The World Bank 2003). CM convenes users and providers to discuss problems around local health service delivery and agree upon actions both groups can take to address these problems.

The CM intervention follows a four-step protocol. First, trained facilitators convene meetings with clinic staff and share scorecards rating local health problems. The scorecard includes five indicators related to maternal and child health (maternal mortality, under-5 mortality, vaccination rate, percentage of births in a health facility, and completion of four antenatal visits). These are constructed from relevant questions in the baseline household survey, and are compared to the district average so as to prompt discussion. Clinic staff are then invited to share their concerns and frustrations with the community. For example, nurses frequently complained that community members do not utilize the clinic, that mothers choose not to give birth in the

clinic, and that parents fail to complete the vaccination course for children.

Second, facilitators convene a meeting of community members excluding the clinic staff, and use the same five indicators to prompt discussion, along with three additional indicators related to user experiences (charging of illegal fees, nurse absenteeism, and staff attitude). Community members are then invited to raise concerns about health outcomes and services. Common complaints include rude behavior from staff and nurses not taking the time to listen carefully to patients' concerns.

Third, interface meetings bring together community members and clinic staff. Facilitators guide a discussion, in which both sides have the opportunity to articulate the complaints and concerns raised in the earlier meetings. The facilitators then assist in the formulation of a "joint action plan" that specifies actions that the clinic staff and community members will take to address issues with health service delivery. For each component of the joint action plan, facilitators work with both sides to specify a time-frame and assign a responsible "point person"; meetings conclude with the signing of the compact by community and clinic representatives.[8] Several of the most common problems cited in the joint action plans relate to utilization, and the actions that users and providers jointly agree to target this outcome. For example, health facility staff are charged with encouraging institutional deliveries, referring and escorting community members to health facilities, discouraging the use of "quacks," and handling patients with a "good attitude." The community agrees to encourage use of clinic services and seek prompt and early treatment. After the meeting, NGOs leave a copy of the compact with the clinic and representatives from each village.

Finally, facilitators held follow-up meetings 1, 3, and 9 months after the initial inter-face meeting to monitor progress on the joint action plans. These meetings included both community members and clinic staff. At the three-month follow-up, the average meeting size is only slightly lower than attendance at the first inter-face meetings.

2.2.2 Non-financial Awards

The Non-financial Awards (NFA) intervention promotes competition among clinics. Clinics in the NFA arm compete to be the best or most-improved clinic in their district; with 85 NFA clinics spread across 4 districts, just under 10 percent of clinics receive an award.[9]

---

[8] The research team randomly selected and monitored half of these initial inter-face meetings. (The other half of clinics in CM were monitored at the three-month follow-up meeting, as noted below.) We find that the vast majority of facilitators adhered to meeting protocols: meetings typically lasted 3–4 hours; the average meeting ranged from 51.9 people in Kenema district to 68.2 in Bombali district and included representatives from the clinic, traditional authorities, and a larger number of community members (with roughly equal representation of men and women).

[9] The average clinic has just over two staff members; this small size ameliorates free-riding problems that might otherwise arise in a competition that awards clinic-wide outcomes, rather than individual effort.

We rank clinics at baseline and endline, using baseline data collected at clinics and from households residing in a catchment area around clinics. Key performance indicators include measures of utilization for ante-natal care, childbirth, and vaccinations; as well as measures related to users' experiences, including absenteeism, staff attitude, and charging fees for services that should be free. Importantly, we do not reveal these indicators publicly to avoid having staff reallocate their effort toward these tasks at the expense of other important tasks.

Although staff do not know what indicators are used in the ranking, we also worry that competing clinics might manipulate their records to exaggerate their performance. To encourage truthful reporting, we inform all clinics (not just those in NFA) that we will audit their registers at baseline and endline and disqualify clinics with fraudulent entries. Each audit involved randomly selecting 30 patients from the clinic register (corresponding to 15 patients per study community) and then visiting these individuals to verify their recorded visit date and purpose. None of the audits uncovered ghost patients or manipulated entries in the clinic register.[10]

In the NFA arm, the competition is announced and extensively advertised at the clinic. As in the CM intervention protocol, facilitators revisit clinics three additional times during the intervention to sustain interest in the competition. Awards are given to the best and the most-improved clinics in each district, with winners announced after the endline survey concludes. Winning clinics receive a plaque or wall clock to display inside the clinic at a public ceremony.

The awards are "non-financial" from the government's perspective, as that they do not involve any monetary compensation. Yet, workers may associate winning with a longer-term financial payoff. For example, workers may have anticipated that being on staff at a wining clinic would facilitate promotions or transfers to attractive locations. We are agnostic about which element of the award, recognition or potential future gains, motivated workers; our aim is to test whether it had any discernible effects on health outcomes or users' experiences.

---

[10] All 254 clinics were told they would be audited at baseline and endline. At baseline, the audit was conducted for all clinics, and clinic staff were also reminded that it would be repeated following the endline survey. During endline we sampled clinic registers to conduct the audits from clinics in all study clinics; however, to reduce data collection costs, we only verified patient details for NFA clinics. The lack of verification in CM and control clinics cannot differentially affect reporting, since the endline data is collected under the common expectation that every clinic will be audited and prior to when endline verification occurred.

3. Design and Methods

3.1 Sampling

3.1.1 Full Sample

We worked in four districts (Bombali, Tonkolili, Bo, and Kenema) that bisect Sierra Leone from north to south (see Figure A.1(a)). These districts include 319 government-run clinics (PHUs) that provide primary and maternal care. We sampled 254 of these clinics, such that all sampled clinics were separated by at least 3 kilometers to minimize spillovers. As a result, the average distance to the next nearest clinic is 10 kilometers. Moreover, in Sierra Leone, residents are "assigned" to clinics (i.e., told by government what clinics to use), which discourages use of more distant clinics. At baseline, the average clinic in our sample had just over 2 staff members present, reported being open 6 days per week, and saw roughly 450 outpatients in the month prior to the survey. Over 80 percent of clinics had walls and roofs in good condition, accessed piped or protected water, and had stocks of basic medications (e.g., oral re-hydration salts and antibiotics); yet, only 10 percent had functional electrical lighting.

We then randomly sample two communities from the catchment area of each clinic (defined as a 3.2 kilometer buffer around the clinic). This generates a sample of 508 communities. At baseline, we randomly sampled 5 households in each of these communities for an extensive household survey (2,540 households). We also randomly sampled an additional 15 individuals in each village and administered a shorter user-feedback survey focused on recent health episodes, service provision, and satisfaction. At endline, we re-surveyed the 5 households that took the baseline household survey. We also randomly selected 5 of the 15 individuals who took the user-feedback survey at baseline and administered the endline household survey, which was revised to incorporate modules from the user-feedback survey. This generates a sample of 10 households per community at endline (5,080 households). The households in our sample are poor: at baseline, 74 percent live in homes with mud floors and wooden walls; 24 percent have no toilet facility and another 58 percent use a pit latrine; only 20 percent own a mobile phone; and 62 percent have no formal education.

3.1.2 Ebola Sub-Sample

Our data on Ebola cases are aggregated to the section-week. Sections are the smallest administrative units in Sierra Leone: the median section is under 40 square kilometers and had fewer than 2,500 residents in the

2004 census (see Figure A.1(c) for a map of section boundaries).[11]

The 254 clinics in the full sample fall into 205 sections. Figure 1(b) maps all sections included in the study with those include in the Ebola sub-sample in dark grey. Of the 205 sections, 45 sections include multiple clinics. In our primary analysis we restrict attention to the 160 sections that contain a single clinic and, thus, a unique treatment assignment. Of the 160, we retain 54 control sections, 46 CM, and 60 NFA. As a robustness check, we also analyze the data using a dose-response model, which uses the proportion of clinics in each section that receive either treatment (see Table E.11).

3.2 Data Collection

3.2.1 Survey Data on Health Clinics, Health Services and Health Outcomes

Baseline data collection took place in September 2011; endline surveys, in May and June of 2013 (see Figure 2 for a timeline). We rely on three survey instruments: first, surveys at each clinic, in which enumerators audited the staffing, cleanliness, drug stocks, and registers of clinics; second, surveys of leaders of each village regarding amenities, relations with the clinic, and community development; and third, household surveys which captured health and economic outcomes and health-seeking behavior among other outcomes.

We filed an analysis plan for the survey data at the AEA RCT registry on March 2017.[12] The analysis plan defines ten outcome families: (1) general utilization, (2) maternal utilization, (3) health outcomes, (4) satisfaction, (5) health service delivery, (6) clinic quality, (7) community development and political engagement, (8) contributions to clinics, (9) water and sanitation, and (10) economic outcomes. The sub-components comprising each outcome family were specified in the analysis plan and are listed in Appendix D.2.

Each outcome family represents a set of variables aggregated using control group-standardized indices per Kling et al. (2007). To create an index of K outcomes, we first reverse outcomes where necessary such that a higher value indicates better outcomes. We then compute $\widetilde{y}_i = \frac{1}{K} \sum^{K} \left(\frac{y_{ik} - \mu_{0k}}{\sigma_{0k}} \right)$, where μ_{0k} and σ_{0k}

---

[11] Our decision to aggregate cases to the section-level stems from two features of our geocoding procedure (see Section E.22). First, many villages in Sierra Leone do not have recorded names; when patients report their residences, they tend to name better-known towns, rather than their village or hamlet. As relevant administrative units, sections typically feature a central or headquarter town. Using a larger administrative unit, we avoid measurement error that arises from attributing patients to larger towns that actually occur in the surrounding villages. Second, our geocoding procedure matches residences to lists of geolocated placenames. When we use smaller geographic units, these often contain few placenames that we can match patients' residences to: 85 percent of census enumeration areas (which are just 7 square kilometers on average) contain one or zero placenames. By contrast, the average section contains 8 geolocated placenames; 94 percent of sections contain more than 1 placename.

[12] AEARCTR-0002085: https://www.socialscienceregistry.org/trials/2085.

are the estimated control-group mean and standard deviation for outcome k in family K. Our estimates for these families represent standard deviation changes relative to the control group.

Following Kling et al. (2007), in case y_{ik} is missing but another sub-component of the family is measured, we impute the mean from the same treatment arm and survey wave. Some sub-components, e.g., those that relate to childbirth, are only defined for a fraction of respondents. As such, we do not impute values when estimating treatment effects for individual sub-components. To demonstrate the imputation is innocuous when looking at effects on families, we follow Kling et al. (2004) and Casey et al. (2012) and aggregate treatment effects across the sub-components of each family using seemingly-unrelated regressions (SUR). These results (reported in Appendix Section D.2 generate nearly identical inferences.

Below we describe each family; Appendix Section B.1 provides additional detail on each family's constituent outcomes.

1. **General utilization** measures the number of episodes in which individuals seek care at a western clinic, including in response to four of the most common health needs relevant for primary health units — childbirth in the past year, antenatal or postnatal care, vaccination, or any illness or injury in the past month — as well as a residual category of any other type of consultation in the past month. The residual category helps generate a comprehensive measure of utilization; we note, however, that regular health check-ups are not common in our setting, so nearly all utilization occurs in response to specific health needs. Utilization of western clinics reflects the decision to seek care at western-style medical facility (overwhelmingly at a government-run clinic given the lack of NGO-run or private clinics), rather than visiting a traditional healer or spiritual leader or avoiding any type of care.

2. **Maternal utilization** is measured among women who gave birth in the year before the endline survey. The family includes two outcomes: an index of the number of times a woman sought antenatal care (ANC) or postnatal care (PNC), and an indicator for whether the woman gave birth in a western-style medical facility.[13]

3. **Health outcomes** are measured at the household-level. The family includes four measures related to child health: under-5 mortality over the past six months; under-5 illnesses over the past month (e.g., malaria or diarrhea); under-2 vaccine completion; and under-5 child wasting, measured using the weight-for-length ratio.[14] The family also includes three other variables: two related to childbirth, maternal mortality over the last six months and problems faced by the mother or newborn within two months of delivery; and one

---

[13] Some outcomes within families are themselves indices. Throughout we use the control-group standardized indices described above (Kling et al. 2007).

[14] We collected data on upper-arm circumference. However, inspection of this variable reveals implausible values due to deviations from our survey protocol: some enumerators wrongly recorded measurements in inches; others, as directed, in centimeters. We cannot discern with certainty which units apply to many observations and, thus, rely on child weight-for-length to measure wasting.

indicator related to general health, whether any household member reports an illness or injury.

4. **Satisfaction** is measured at the household-level. The family includes three outcomes measured on a four-point Likert scale from "Very Unsatisfied" to "Very Satistfied": the respondent's satisfaction with their family's health; satisfaction with public health workers (i.e., clinic staff); and, among households with at least one member utilizing a western-style clinic in the last year (approximately half), satisfaction with the care they received.[15] Among these households with members utilizing the clinic in the last year, we ask whether they would return to the clinic for a future medical need. The last two satisfaction outcomes are asked across the different types of health episodes, so we average responses across individuals in a household when multiple episodes are reported.

5. **Clinic quality** includes three clinic-level outcomes. First, we construct an index of clinic service provision that aggregates measures related to organization (e.g., medicines sorted by expiration date and stored in a safe location), the types and frequency of services offered (e.g., family planning), number of staff on duty, and hours clinics are open. Second, we measure the proportion of staff who are aware of the 2010 policy that removed user fees for maternal and under-5 services. Finally, we measure employee satisfaction. The services offered and employee satisfaction are reported in the clinic survey; other measures are based on enumerators' observations.

6. **Health service delivery** is measured among individuals who experience a health episode in the month before the endline survey (for childbirth episodes recall is over the past six months). The family includes outcomes derived from the household survey, including staff absenteeism and wait times, problems with clinic facilities or staff, satisfaction with services, staff attitude, drug availability, and fees paid. Per our analysis plan, two satisfaction outcomes appear in this family and the satisfaction family (measured at the individual- and household-level respectively). We verify that this redundancy does not affect our family-level result for health service delivery.

7. **Contributions to clinics** is measured at the community-level. The family includes two outcomes. The first outcome, derived from the survey of village leaders, captures whether the community convened meetings about the clinic and whether it contributed labor to the upkeep of the clinic or well-being of staff (e.g., helping to plant a garden for nurses). The second outcome incorporates responses from clinic staff about whether the community made such contributions or had disputes with clinic staff.

8. **Community development and political engagement (CDPE)** is measured at the community-level. The family includes outcomes related to community members' participation in meetings in the last six months, contributions to local development projects over the last year, their self-reported ability to address problems collectively over the last year, and turnout for the local and national elections in November 2012.

---

[15] As with general utilization, satisfaction with care is asked of individuals who attend the clinic for childbirth in the past year, antenatal or postnatal care, vaccination, any illness or injury in the past month or any other type of consultation in the past month.

9. **Water and sanitation** is measured at the household-level and includes three outcomes: an index that tracks households' access to potable water and toilet facilities; an index that measures public water and toilet facilities in each community; and an index of questions related to households' satisfaction (measured on the four-point Likert scale) with water, sanitation, and cleanliness in their community.

10. **Economic outcomes** is measured at the household-level and comprises four outcomes: indices of physical assets, agricultural assets (e.g., livestock, farm tools), and dwelling materials; as well as an index capturing total consumption expenditure over the last month.

Families 1–4 (general utilization, maternal utilization, health outcomes, and satisfaction) convey the main effects and mechanisms for our medium-run results. Families 5–8 (clinic quality, health service delivery, contributions to clinic, and CDPE) represent other plausible but unsupported mechanisms. Finally, families 9 and 10 (water and sanitation and economic outcomes) are additional outcomes presented in the appendix.

3.2.2   Ebola Case Data

We rely on the Epi Info Viral Hemorrhagic Fever (VHF) database, which was the primary data management system used for case and contact tracing during the outbreak. The Ministry of Health and Sanitation, with support from the CDC implemented and maintained the VHF database through the end of the epidemic, and McNamara et al. (2016, 39) describe it as "the most comprehensive epidemiologic and laboratory data on Ebola cases available in Sierra Leone." As noted above, the VHF cannot be used to measure the underlying incidence of Ebola; rather, it reflects reported cases — a particularly important outcome for stopping contagion and containing the epidemic (Enserink 2014).

We use a de-identified version of the VHF database (where patient names and characteristics have been redacted).[16] While the data were compiled at the National Ebola Response Center in Freetown, data entry occurred at the district-level. Surveillance officers employed (even prior to the crisis) by the Ministry of Health and Sanitation (MoHS) oversaw teams of case investigators charged with following up on suspected cases. Investigators learned about cases from communities, phone calls, active surveillance (e.g., contact tracing), and through walk-ins at health centers (Owada et al. 2016, 2). For each suspected case, they completed a case investigation form (CIF), which included demographic (including district, chiefdom, and village) and health information. A copy of the CIF accompanied blood samples or swabs from corpses, so that lab results could be linked back to cases to confirm or rule out an infection (McNamara et al. 2016, 37). Completed CIFs were brought back to District Ebola Response Centers and entered by data managers in the

---

[16] The Njala University Ebola Museum and Archive facilitated access to this database.

14

local VHF database. Each observation in our data represents one of these CIFs. Data from the districts were periodically transmitted to the National Ebola Response Center.

The VHF database includes four different types of cases. Two classifications result from testing: **negative** cases where Ebola has been ruled out; and **confirmed** cases per lab results. In total, the VHF contains data on 8358 confirmed cases. The average patient (reported in Fang et al. (2016)) was 30 years old; and over 70% were above the age of 18. There are two residual categories that are never confirmed with lab tests: **suspected** cases display Ebola symptoms and/or have had contact with potentially infected individuals or animals;[17] and **probable** cases meet the criteria for a suspected case and were either screened by a clinician or died and have an epidemiological link to a confirmed case. Given our interest in reporting, our primary outcome is total cases — the sum across these four case types.

We geocode cases using information on individuals' residences included in the VHF database (typically, district, chiefdom, and village or parish). The full geocoding protocol is described in Section E.22. While this geocoding process introduces measurement error, we expect this to be uncorrelated with our randomized treatments and, if anything, result in a loss of efficiency. Consistent with this claim, we test whether treated sections have more or longer placenames, or placenames that are more likely to contain spaces (i.e., two word names), and find no substantial or significant differences (see Table E.27). To check the accuracy of our geocoding protocol, we look at the average number of confirmed cases in sections that hosted major Ebola-specific treatment facilities: Ebola Treatment Units (ETUs), Ebola Holding Centers (EHCs), or Community Care Centers (CCCs). We find that sections with ETUs, the largest facilities, reported 484 cases on average; sections without any of the three facilities averaged just 25.2 total cases (see Table E.1).

Our main dependent variable is the count of total cases aggregated to the section-week. We use the date when a case is first entered in the VHF database to determine the week. To demonstrate robustness, we employ a number of alternative specifications, using the inverse hyperbolic sine transformation of the count and the logged count, as well as estimating linear probability and Poisson models (see Tables 3 and E.8).

---

[17] Suspected cases include (1) the onset of high fever and contact with a suspected, probable, or confirmed individual or a dead or sick animal; (2) the onset of high fever and at least three of the following symptoms: headaches, vomiting, anorexia/loss of appetite, diarrhea, lethargy, stomach pain, aching muscles or joints, difficulty swallowing, breathing difficulties, or hiccup; any person with inexplicable bleeding; or any sudden, inexplicable death. Suspected and probable cases may have died prior to a lab sample being collected; alternatively, administrative issues may have led to tests being missed or not entered into the VHF.

Figure 1: Mapping of Ebola Cases and Sample

(a) Heat Map of VHF Cases by Section   **(b)** Sections Used in Ebola Analysis
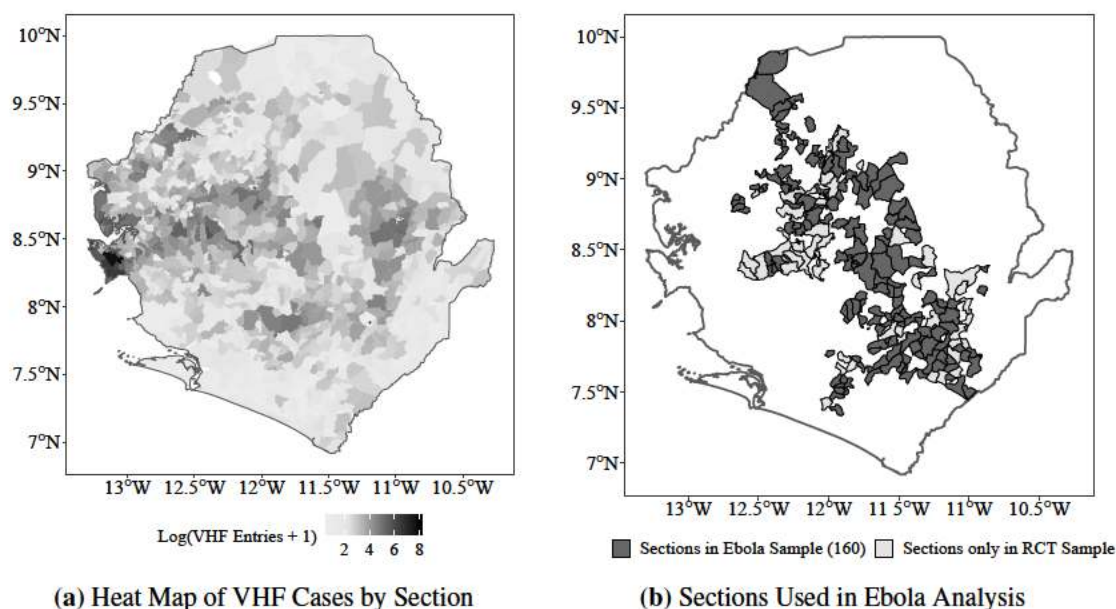
Figure 1(a): the number of entries (logged) by section in the Viral Hemorrhagic Fever (VHF) database maintained by the CDC during the Ebola crisis. Figure 1(b): map of all sections that contain clinics that were part of the original randomized experiment. Sections in light gray are excluded, because they contain more than one clinic from the original RCT.

3.3   Randomization

3.3.1   Matching and Blocking

We grouped the 254 clinics in our sample into matched triplets. Clinics in a triplet fall within the same district and exhibit similar levels of utilization and performance at baseline.[18] Within each triplet, we then randomize clinics into control, community monitoring (CM), or non-financial awards (NFA). This results in 84 clinics assigned to control, 85 to CM, and 85 to NFA. Figure 2 illustrates this allocation of treatment and the timing of randomization relative to data collection. We include matched-triplet fixed effects in our reduced-form specifications to account for the blocked randomization.

---

[18] We exactly match clinics by district and type (MCHP or CHP) and then select matches based on the Mahalanobis distance between eight indicators specified by the Ministry of Health: completion of first-year vaccinations, institutional deliveries, completion of fourth antenatal care visit, charging of fees for maternal and under-5 services, nurse absenteeism, staff attitude, maternal mortality, and under-5 mortality.

Figure 2: Consort Diagram
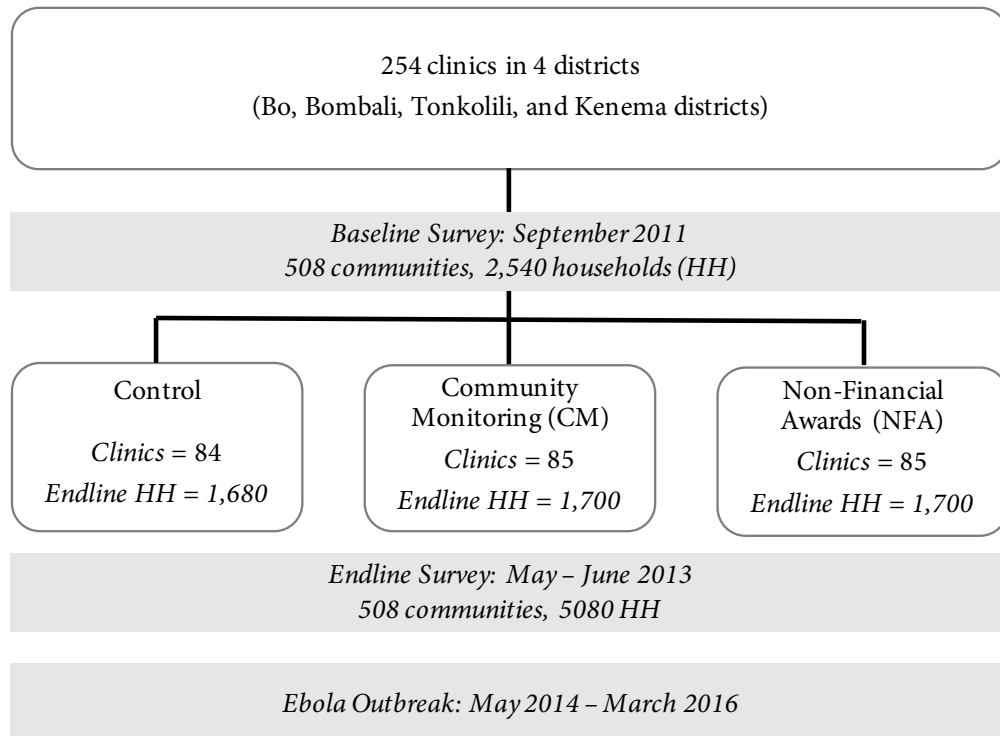
Figure 2: the samples and timing associated baseline and endline surveys, the randomization, and the Ebola crisis. An initial end to the outbreak was declared in November 2015; however, a few additional cases emerged and the country was finally deemed "Ebola free" in March 2016.

3.4   Integrity of the Experiment

Table C.1 reports balance across pre-specified covariates. Most notably, we find that the number of injuries or illnesses reported is lower in both treatment arms relative to control, household size is slightly smaller, and there are fewer households reporting a recent childbirth in CM. If anything, we expect such imbalances make it harder to find effects on general and maternal utilization. We also find that NFA communities have better cell phone coverage and a higher level of educational attainment, are less likely to belong to the Temne ethic group and are less likely to believe what a doctor told them. As a robustness check, we include unbalanced variables at baseline as additional controls.

We report manipulation checks in Tables C.2 and C.3 based on survey responses. Over 85 percent of CM communities report a meeting held by one of our implementing partners to discuss working with the clinic to improve health service delivery; the average CM community reports 2.5 such meetings. Around 44 percent of control communities report meetings as well. This does not reflect non-compliance by our implementing partners; rather, community meetings unrelated to our interventions are not uncommon across

these districts and were likely to have been mistaken for interface meetings. Nearly 95 percent of staff in NFA clinics have heard of the NFA intervention; conditional on knowing about the program, 84 percent report participating compared to 48 percent among control villages. Staff in CM and control clinics also appear to have heard about the NFA program, albeit at lower rates. Conditional on having heard about NFA, 13 of 84 control clinics report participating. Again, this likely reflects misperception and not implementation failures or contamination: control clinics cannot opt into the NFA competition, which requires the research team to rank each participating clinic. Moreover, the research staff did not detect any deviations from treatment assignment when monitoring implementation of the NFA treatment.[19] If clinic staff in control clinics are motivated by the mistaken belief that they are eligible for an award, this would attenuate our treatment effects.

3.5   Specifications

3.5.1   Survey Outcomes

The analysis plan specifies outcome variables, the construction of indices, and the specifications we estimate; and we flag and explain any subsequent deviations in Appendix Section B.1.

Our baseline survey included a smaller sample of households. Moreover, some outcomes are only defined for individuals who recently experienced health episodes. Some households that experienced health episodes at endline may not have also experienced episodes during baseline. For both reasons, controlling for a household's baseline outcome would reduce the size and representativeness of our sample. We instead employ an ANCOVA-type model that controls for the village-level average at baseline, estimating:

$$y_{ivc,EL} = \alpha_b + \beta^{\text{CM}} \mathbb{1}(CM)_c + \beta^{\text{NFA}} \mathbb{1}(NFA)_c + \delta \overline{Y}_{vc,BL} + \varepsilon_{ivc,EL} \tag{1}$$

where $y_{ivc,EL}$ is the outcome of household (or individual) i in village v in clinic catchment c at endline (*EL*). (Models for community-level outcomes omit the i subscript.) α_b represents the matched-triplet fixed effects. Treatment status, which is randomized across clinics, is denoted by the indicator variables $\mathbb{1}(CM)_c$ and $\mathbb{1}(NFA)_c$. $\overline{Y}_{vc,BL}$ is the village-level average at baseline. When y is a sub-component of an outcome family, we still use the family-level outcome to compute this baseline average.[20] We cluster our standard errors on clinic, the unit of randomization. We also estimate a variant of Equation 1 in which we pool the CM and

---

[19] The NFA treatment involved 4 rounds of meetings. During the second round, we randomly selected half of all clinics and sent an enumerator to monitor the meeting.

[20] This decision is motivated by two features of our data: first, some sub-components are only measured at endline; second, for some sub-components and villages we have no data to compute the average (e.g., if there were no recent births). To improve precision through the inclusion of a prognostic pre-treatment covariate, we include the average family-level outcome. This represents a slight deviation from the analysis plan, but does not affect affect any of our conclusions.

NFA treatments into one pooled treatment indicator.

When analyzing data at the clinic-level, we drop the indexes for households and villages, estimating:

$$y_{c,EL} = \alpha_b + \beta^{CM}\mathbb{1}(CM)_c + \beta^{NFA}\mathbb{1}(NFA)_c + \delta\overline{Y}_{c,BL} + \varepsilon_{c,EL} \qquad (2)$$

These models include a single observations for each clinic, so there is no need to cluster our standard errors.

In addition to conventional standard errors, we report q-values that control for the proportion of incorrectly rejected null hypotheses (Anderson 2008). Specifically, we control for the false discovery rate within treatment arm (1) across outcome families; and (2) across sub-components within each family.[21]

As noted above, we follow Kling et al. (2007) to construct outcome families for the medium-run results. If an individual only has data for some sub-components within a family, this procedure involves imputing those missing sub-components using the mean from that individual's treatment arm. This leads to high rates of imputation when a family includes sub-components that are not well-defined for a large share of observations (e.g., complications with childbirth can only be measured for mothers). To confirm that such imputation does not influence our results, we follow Kling et al. (2007) and Casey et al. (2012) and employ seemingly unrelated regressions (SUR). In short, we estimate treatment effects on each sub-component of a family without any imputation and then average the coefficients across these stacked regressions, using the delta method to compute standard errors. The tables in Appendix Section D.2 include results for our pre-specified families using SUR, as well as the results from the stacked sub-component regressions.

3.5.2 Ebola Case Data

We assess the impact of the CM and NFA interventions on reported cases for the 160 sections in the Ebola sample (described in Section 3.1.2). We observe counts of reported cases in each section in every week from 10 August 2014 to 18 October 2015. However, we restrict attention to the period from September through April 2015, when Ebola transmission was a real threat in our study area; less than 1 percent of confirmed cases (3 in total) occur in the seven months between May and October 2015.[22]

---

[21] In the analysis plan, we specified controlling for the false discovery rate (FDR) only across some families denoted primary families and within those families, only across some outcomes. However, since we examine all outcomes, and several key to our overall account (e.g., satisfaction) were designated such that they would not have been accounted for in the FDR adjustment, we deviate from the analysis plan in the conservative direction by instead, correcting across all outcome families, and within each family, across all composite outcomes.

[22] In Table E.9 we extend the panel back to August 2014 and replicate our primary results from Table 3.

Using this data, we estimate:

$$y_{st} = \alpha_b + \delta_t + \gamma^{CM} \mathbb{1}(CM)_s + \gamma^{NFA} \mathbb{1}(NFA)_s + \eta_{st} \qquad (3)$$

where α_b again represents the matched-triplet fixed effects; δ_t are week fixed effects; $s \in \{1, 2, \ldots, 160\}$ indexes sections; and $t \in \{1, 2, \ldots, 34\}$ indexes weeks. For panel models, we cluster our standard errors at the section-level, which in the Ebola sample coincides with the clinic, the level of randomization.[23]

We amend Equation 3 to detect spillovers within our study sample — namely, the reallocation of patients from control to treated sections (or vice versa). Specifically, we interact our treatment indicators with covariates that, in the presence of such spillovers, should moderate our treatment effects (e.g., distance between sections, connections via roads).

4. Results

4.1 Medium-term Outcomes

We present the main medium-run survey results in Table 1, which shows estimated effects on clinic utilization, health outcomes, and patient satisfaction.[24] Our tables follow a common format. Column 1 provides the control mean and standard deviation, which are zero and one exactly when looking at family-level mean-effects indices. Column 2 presents the average treatment effect (in standard-deviation units) when pooling the treatment arms in a comparison with control. Columns 3 and 4 separately estimate the average treatment effects for CM and NFA, respectively. Column 5 shows the difference between the average treatment effect in CM and NFA. Column 6 provides the F-test for the joint null hypothesis of no effect from either treatment. Finally, Column 7 gives the sample size used for each regression.

Table 1 shows uniformly positive treatment effects across families for CM. Specifically, the general utilization of clinics increases by 0.11 standard deviation units when we pool the treatments, and the separate effects are similar for CM and NFA. Individuals in the control group recently used a western clinic for

---

[23] When we collapse the data over time and estimate cross-sectional models, we omit the week fixed effects and t subscripts. As treatment assignment occurs below the section-level, we do not cluster our standard errors in the cross-sectional models.

[24] To gauge how individual indicators contribute to our family-level effects, Appendix Tables D.1–D.10 present treatment effects for each of the individual indicators. Tables D.11–D.19 repeat these analyses using the z-scored (i.e., control group-standardized) versions of the indicators. As discussed above, our primary approach to aggregating across individual indicators is to construct families of outcomes (i.e., mean-effects indices) per Kling et al. (2007). As a robustness check, we also use SUR to average treatment effects across indicators, and the results from this approach are presented in the top rows of the tables in Appendix Section D. Overall, our results are very similar using either method of aggregation; some coefficients attenuate under SUR, notably satisfaction and CDPE.

close to one (0.96) health episode (see Table D.1); the treatments increase utilization by about 5 percentage points.[25]

Maternal utilization — the use of clinics for ante-natal care, birthing, and post-natal visits — can only be measured among the 888 women who gave birth in the year preceding the endline survey. In this sample, we find that maternal utilization increases by 0.18 standard deviation units in CM, but there is no equivalent effect for NFA (or when we pool the treatments). As shown in Appendix Table D.2, increased maternal utilization in CM is driven by more deliveries at western-style medical facilities: the probability of giving birth in a western facility is 0.83 in control areas; CM boosts this rate by 0.09 percentage points, an 11 percent increase.

The health outcomes family also increases by 0.15 standard deviation units with CM. To a large extent this is driven by significant improvements in child health. As shown in Table D.3, the likelihood that a child under 5 dies in CM falls by 0.015 relative to the control mean of 0.039 — a 38.4 percent effect. In addition, child weight-for-length increases by 0.16 z-score and is significant at the 10% level, though this individual indicator loses significance after FDR adjustments. It is worth noting that the magnitude of these effects are qualitatively similar to those uncovered by Björkman and Svensson (2009) in their study of community monitoring in Uganda. Finally, the effect size for vaccine completion is also large, corresponding to a 10 percent increase in vaccine completion, though the indicator does not reach significance at conventional levels.

---

[25] Our utilization measure includes the use of private or NGO-run clinics, in addition to the government-run clinics (PHUs) targeted by the treatments. These non-governmental providers constitute a small share of utilization (3 percent). When we focus attention on the utilization of just government-run clinics in Table D.21, our treatment effects increase to 7.3 and 6 percent for CM and NFA respectively, verifying that the interventions boosted utilization in the targeted clinics.

Table 1: Utilization, Satisfaction, and Health Outcomes

	(1) Control Mean	(2) Pooled	(3) CM	(4) NFA	(5) Difference	(6) Joint F-test (p)	(7) N
General utilization	0.000 (1.000)	0.112 (0.031)*** [0.005]***	0.126 (0.034)*** [0.003]***	0.099 (0.037)*** [0.032]**	0.026 (0.033)	7.054 (0.001)***	4496
Maternal utilization	0.000 (1.000)	0.061 (0.064) [0.327]	0.175 (0.077)** [0.068]*	−0.043 (0.076) [0.548]	0.218 (0.081)***	4.128 (0.017)**	888
Health outcomes	0.000 (1.000)	0.053 (0.051) [0.327]	0.146 (0.056)*** [0.045]**	−0.037 (0.059) [0.548]	0.184 (0.056)***	6.318 (0.002)***	5053
Satisfaction	0.000 (1.000)	0.101 (0.042)** [0.038]**	0.091 (0.049)* [0.095]*	0.109 (0.049)** [0.048]**	−0.018 (0.048)	2.876 (0.058)*	5052

Notes: Treatment effects are estimated using Missing-Indicator ANCOVA, controlling for the community-level average of the outcome family index at baseline and matching-triplet fixed effects. Robust standard errors, clustered by clinic, are shown in parentheses. Multiple-inference corrected q-values that adjust for the false discovery rate within treatment arm are shown in square brackets. Column (5) presents an F-test for the joint significance of CM and NFA, with the associated p-value in parentheses. Significance: * is significant at the 10% level, ** is significant at the 5% level and *** is significant at the 1% level.

Under CM, women use clinics more during pregnancy and childbirth, and their children's health and survival improves substantially. By contrast, the NFA intervention does not appear to significantly affect either maternal utilization or health outcomes. The scorecards used in the CM intervention focus attention on child and maternal health — these are the only metrics discussed in clinic meeting and a majority of metrics presented at community meetings. It is possible that this emphasis leads to greater improvements for these families in CM, as compared to NFA.

The final row of Table 1 presents impacts on patient satisfaction. Across both treatment arms, patient satisfaction increases by about 0.10 standard deviations, largely driven by increases in respondents' satisfaction with their own health and health workers (see Table D.4). These effects are consistent with the idea that the interventions led to improvements in the quality of care provided by health workers in ways that are difficult to capture in survey measures.[26] For example, improvements in quality could include the extent to which health providers listen to patients describing symptoms, or the effort they expend on diagnosis and selecting and explaining an appropriate plan of care, which would translate into greater satisfaction with their performance.

The results in Table 1 show that these interventions, which were designed to make frontline health staff more accountable to users, can generate substantial improvements in the quality of care that patients anticipate when visiting a clinic. Both interventions increase general utilization of clinics, a behavior that reveals would-be patients' increased confidence in clinics, as well as patients' reported satisfaction with their health and with health workers.

We now turn to other families of outcomes that might help explain increased utilization and, in CM, improved health outcomes.[27] Specifically, in Table 2, we examine effects on inputs and resources available at clinics (clinic quality), as well as the types and quantity of services provided (health service delivery). We also examine community support for the clinic and community development and political engagement (CDPE). A supply-side response — more resources at clinics or a larger menu of services — or groundswell of community support could also draw would-be patients into clinics and improve their outcomes.

Yet, we do not observe effects on the resources or services available at the clinic: the clinic quality and health service delivery families do not change significantly in response to either treatment (or jointly). This is perhaps unsurprising given that the CM and NFA interventions do not facilitate clinics' access to supplies or training; the Government's interest in the evaluation is understanding how to extract greater effort from health workers under existing budget and logistical constraints. None of the sub-components of clinic quality significantly improve (see Table D.5). And the coefficients are essentially zero or negative for those sub-

---

[26] Patient-provider interactions are rarely directly observed; for a rare exception see (Das et al. 2016).

[27] Spillovers do not provide a plausible explanation: our utilization measure is based on household surveys, and not clinic registers. If our endline respondents traveled to treated clinics for care this would attenuate our medium-run results, as it would appear to increase utilization among households living in the catchment areas of control clinics.

components of health service delivery related to supply constraints (e.g., availability of drugs, staff not present (see Table D.6).[28] We do, however, find substantively meaningful effects on some sub-components of health service delivery that relate to patients' experiences. For example, we see a 59 percent reduction in unpleasant staff behavior in NFA areas, though the effect is not statistically significant. The coefficient for staff attitude also suggests improvements, though it loses significance after the FDR adjustments.[29] While there are no significant changes in the services offered at clinics, we do see indications of more positive interactions between patients and staff. This could help to explain the improvements in satisfaction, particularly in NFA where we do not find changes in health outcomes.[30]

We also find no increase in community support: community members did not spend more time or resources on the clinic or its staff. We do observe improvements in overall community development and political engagement (CDPE). Yet, the sub-components driving this effect seem unlikely to affect health-seeking behavior (see Table D.8). For example, there are no significant effects on the community coming together to address problems collectively. Instead, the effects are largely driven by contributions to local development projects in both CM and NFA communities and small (<0.5 percent) increases in voter turnout. [31] We also observe improvements in our water and sanitation family when we pool the treatments, driven by NFA communities (see Appendix Table D.9).[32] Finally, we find weak effects on economic outcomes (see Appendix Table D.10), suggesting that the interventions did not materially affect households in treated communities.

As a robustness check, we control for imbalanced baseline covariates in Table D.23. Only the effects on community development and political engagement attenuate. Our ANCOVA specification controls for the baseline value of each family and thus addresses the direct effects of any baseline imbalance in that outcome.

---

[28] We observe a positive effect of NFA on absenteeism in Table D.6. This is likely an artefact of how we specify this measure: we ask respondents "of all the times you visited the clinic in the past month, did you ever find there were no staff present?" An obvious drawback is that an individual who visits the clinic more frequently has more opportunities to find staff absent. Given the treatment effects on general utilization that we report above, it seems likely that such post-treatment bias pushes towards a positive relationship between the interventions and this measure of absenteeism. Fortunately, we also ask whether respondents found staff absent during their last visit to the clinic. Table D.22 shows precise null effects on this outcome. This comparison suggests that the positive effect on absenteeism in Table D.6 likely reflects increases in utilization, not rates of absenteeism.

[29] There is a small (about 1.5 percent) and marginally significant increase in whether people would return to the clinic in the future. Note, however, the ceiling effects may limit our ability to detect improvements using this measure: nearly all (97 percent) patients in control areas report that they would return.

[30] We observe null effects on health service delivery despite including the "satisfaction with care" and "would return to clinic" variables, which are also sub-components of our satisfaction family. This reflects our original analysis plan; however, we verify that removing these two indicators does not meaningfully alter the null effect on this family. These results are available upon request.

[31] These small effects could indicate that people changed their beliefs about the government and hence, political participation. Note that the control group turnout is already approaching the maximum at 97 percent.

[32] These effects reflect greater access to mechanical wells in these communities. The greater availability of wells is consistent with greater community participation in development projects, which would provide the labor and resources needed to put wells into place.

The positive effects on clinic utilization, health, and satisfaction are not driven by "top-down" improvements in the supply of health services or by greater community contributions to clinics. Rather people utilize the clinic to a greater degree, report greater satisfaction (potentially due to improvements in the quality of service delivery such as more positive interactions with staff), and, in CM, see improvements in child health outcomes.

Table 2: Supply-side Measures and Community Support

	(1) Control Mean	(2) Pooled	(3) CM	(4) NFA	(5) Difference	(6) Joint F-test (p)	(7) N
Clinic quality	0.000	0.104	−0.004	0.213	−0.216	0.929	254
	(1.000)	(0.149)	(0.175)	(0.176)	(0.184)	(0.397)	
		[0.395]	[0.649]	[0.237]			
Health service delivery	0.000	0.039	0.070	0.027	0.043	0.507	2877
	(1.000)	(0.059)	(0.071)	(0.062)	(0.059)	(0.603)	
		[0.395]	[0.266]	[0.583]			
Community support	0.000	0.007	0.027	−0.013	0.040	0.062	508
	(1.000)	(0.095)	(0.112)	(0.109)	(0.116)	(0.940)	
		[0.891]	[0.564]	[0.619]			
CDPE	0.000	0.231	0.202	0.261	−0.059	3.849	508
	(1.000)	(0.085)***	(0.102)**	(0.101)**	(0.110)	(0.023)**	
		[0.034]**	[0.095]*	[0.032]**			

Notes: Treatment effects are estimated using ANCOVA, controlling for the community-level average of the outcome family index at baseline and matching-triplet fixed effects. Robust standard errors, clustered by clinic, are shown in parentheses. Multiple-inference corrected q-values that adjust for the false discovery rate within treatment arm are shown in square brackets. Column (5) presents an F-test for the joint significance of CM and NFA, with the associated p-value in parentheses. Significance: * is significant at the 10% level, ** is significant at the 5% level and *** is significant at the 1% level.

4.2   Long-run Effects on Ebola Outcomes

Roughly one year after our endline survey, the first confirmed Ebola case was recorded in Sierra Leone. We turn now to examining the longer-run effects that the interventions had on reporting and mortality during the ensuing epidemic.

The treatment effects on reported Ebola cases are apparent in Figure 3: the left panel presents the sum of total reported cases in each week by treatment arm; the right panel is the cumulative count of cases during our study period. Between September 2014 and May 2015, we count 515 total cases in control sections; yet in sections with clinics receiving the CM and NFA interventions 735 and 795 cases are reported, respectively. This difference is even more striking for confirmed cases: only 21 confirmed cases are reported in the 54 control sections, while 248 are reported in the treated sections (see Figure E.2).

Figure 3: Total Cases by Treatment

(a) Weekly Counts   (b) Cumulative Counts
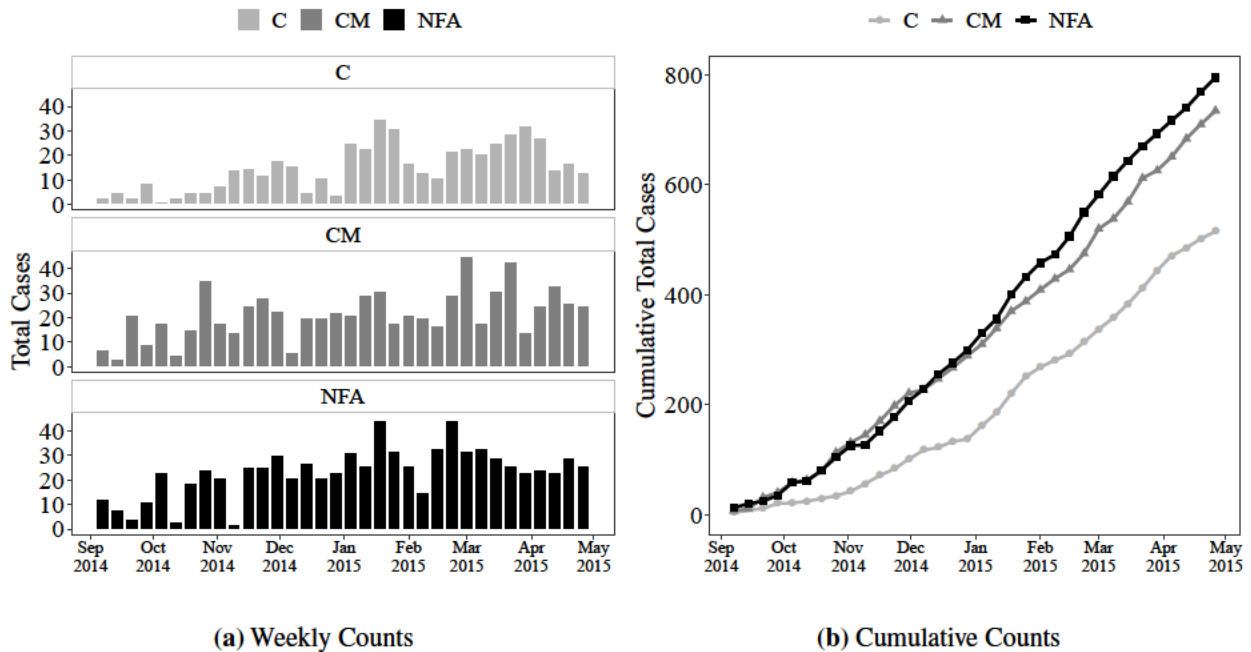
Figure 3(a) plots the time series of total cases by week; bars represent the raw counts. We use the date that the case was first saved in the VHF. Figure 3(b) graphs the cumulative count of total cases by treatment group.

We report regression results using Equation 3 in Table 3. In the top panel, the outcomes are the raw counts of total, confirmed, and negative cases. The pooled effect implies a 62 percent increase in the average number of total cases. The effect is smaller and less precisely estimated for NFA ($p = 0.13$), which is consistent with our medium-run results, which also show smaller effects on utilization for NFA clinics, especially in the Ebola sample (see first row of Table 5). Nonetheless, we cannot reject the null hypothesis

27

that the treatments have equivalent effects. To both improve patient survival and contain the epidemic, it is particularly important that infected patients report. We find large increases in the average number of confirmed cases reporting in treated sections: for every confirmed case in control, we count five confirmed cases in treated sections.

Table 3: Reported Ebola Cases

	Control Mean	Pooled	CM	NFA	Difference	N
Ebola Cases						
Total	0.281	0.173	0.204	0.148	0.055	5,440
	(0.727)	(0.084)**	(0.117)*	(0.099)	(0.133)	
Confirmed	0.011	0.059	0.086	0.039	0.047	5,440
	(0.129)	(0.024)**	(0.038)**	(0.025)	(0.041)	
Negative	0.238	0.1	0.079	0.115	-0.036	5,440
	(0.648)	(0.061)	(0.077)	(0.075)	(0.093)	
IHS(Ebola Cases)						
Total	1.015	0.057	0.065	0.05	0.014	5,440
	(0.309)	(0.029)**	(0.038)*	(0.034)	(0.044)	
Confirmed	0.887	0.02	0.024	0.016	0.007	5,440
	(0.064)	(0.007)***	(0.01)**	(0.008)**	(0.012)	
Negative	0.998	0.039	0.034	0.043	-0.008	5,440
	(0.284)	(0.023)*	(0.03)	(0.029)	(0.035)	

Notes: The sample includes 160 sections over 34 weeks. Standard errors clustered on section shown in parentheses. Treatment effects estimated using OLS including matching-triplet and week fixed effects. Difference column reports the difference between the CM and NFA coefficients; the standard error is computed using the delta method. The bottom panel employs the inverse hyperbolic sine transformation ($\text{IHS}(y) = \log(y + \sqrt{(1 + y^2)})$). Significance: * $p < 0.10$, ** $p < 0.05$, and *** $p < 0.01$.

We take a number of steps to assess robustness. First, we re-estimate our pooled effect dropping one matched triplet at a time (Figure E.4), dropping each possible pair of matched triplets (Figure E.5), or dropping each week (Figure E.6). Second, we estimate the effects by month to assess whether our results are driven by a particular moment in the crisis. The pooled coefficient is positive across every month, and we find significant effects in October, December, February and April (Table E.3). (We find large and significant effects for CM in October 2014 and April 2015; for NFA, in October and December 2014.) The spread of these effects across various months throughout our sample period verifies that the effects are not driven by any particular period. Third, we conduct a placebo test where we substitute the nearest out-of-sample neighbor for each section. We find no significant effects in Table E.12, alleviating concerns that our treated sections are spatially clustered in areas where reporting is higher for reasons unrelated to treatment (e.g., exposure). Finally, we adopt a number of alternative specifications to handle our count data. In the bottom panel of Table 3, we use the inverse hyperbolic sine transformation of the counts and find effects of similar magnitude; the effect of NFA on confirmed cases becomes significant in this specification. We expand upon this in Table E.8, which presents estimates using a linear probability model, the logged count (adding

one to avoid dropping section-weeks with no cases), and a Poisson count model.[33] In the Poisson count model, NFA has significant effects on both confirmed and negative reported cases; the p-value for NFA when analyzing total cases just misses a conventional threshold at 0.104.

Probable and suspected cases (which constitute 1 and 6.5 percent of total cases) are included in our count of total reported cases. However, these cases often do not involve reporting by individuals; their ambiguous status reflects the absence of a definitive lab test (e.g., confirmed or negative). These cases include, for example, deceased individuals with Ebola symptoms. We separately analyze these cases in Table E.6 and find insignificant and negligible treatment effects.[34] Our results in Table 3 are driven by increases in the number of patients that report and receive testing (see Table E.7 which subtracts probable and suspected cases from total cases).

We next look at whether the treatments had any effect on patient deaths.[35] We regress the number of deaths in each section-week on the total number of cases reported in the current and previous week and the interaction of that caseload with treatment. We opt for the caseload over the current and previous week, as Ebola deaths typically occur 6 to 16 days after symptom onset. To ease interpretation, in Table 4 we predict the number of deaths in control and treated sections for a two-week caseload of 2, 5, and 10 cases (see Table E.4 for corresponding regression results). We find significant reductions in mortality: in control sections, we estimate 1 patient death for every 4 cases; that drops to 1 death for every 7 cases in treated sections — a reduction driven by CM, where there is just over 1 patient death for every 10 cases. One may worry that patients in control simply waited longer to report and, thus, presented with a higher risk of mortality. Yet, we show in Table E.5 that treatment does not reduce the number of days between symptom onset and reporting. Analogous to our medium-run effects, improvements in health outcomes during the Ebola crisis concentrate in areas receiving the CM intervention.

These conditional-on-positives estimates will be confounded if treatment changes the composition of patients (e.g., their co-morbidities). The increased number of confirmed patients in treated sections should, if anything, attenuate these results. Yet, despite more infected cases reporting, our findings suggest that patients in treated sections — especially those in CM — enjoyed higher survival rates. Qualitative accounts during the crisis suggest that trust in clinic staff encouraged patients to truthfully report their symptoms and adhere to advice regarding treatment. Raven et al. (2018) also report that health worker morale led to more effective care, especially when treating children.[36] Consistent with the medium-run results, these findings suggest that the accountability interventions improved the care received by patients, even under

---

[33] We also collapse the data and estimate cross-sectional models (Table E.10). Our coefficients are of the same magnitude, but we lose power and precision; the Poisson count models remain highly significant with only 160 observations.

[34] CM has a negligible positive effect on probable and suspected cases; NFA, a negligible negative effect. The resulting difference is small in magnitude — 9 probable and suspected cases cases spread over 106 sections and 34 weeks — but is significant at the 10-percent level.

[35] Sierra Leone lacks vital statistics data, so we can only analyze mortality for cases in the VHF.

[36] See https://blogs.unicef.org/blog/ebola-in-sierra-leone-the-dont-touch-rule/.

Table 4: Patient Deaths

Total Cases in Last 2 Weeks	Predicted Deaths in Control	Predicted Deaths in Pooled	Difference
2 reported cases	0.49	0.36	0.13
	(0.04)	(0.05)	(0.06)**
5 reported cases	1.23	0.80	0.43
	(0.11)	(0.17)	(0.19)**
10 reported cases	2.45	1.53	0.92
	(0.21)	(0.36)	(0.40)**

Notes: Standard errors clustered on section. Treatment effects estimated using OLS including matching-triplet and week fixed effects. Predicted deaths based on estimates in Table E.4 model 1. Significance: *p < 0.10, ** p < 0.05, and *** p < 0.01.

crisis conditions.

4.2.1 Additional Checks within Ebola Sample

Table E.13 reports balance checks for the 160 sections in our Ebola sample. We find negligible differences along most baseline measures relative to levels in control; the imbalance observed in the full sample carries over to this subset. As a robustness check, we aggregate unbalanced baseline indicators to the section-level and includes these as controls (Table E.26). In Table E.20, we also look at whether treated sections are more exposed to the epidemic due to their proximity to index cases in Guinea or Sierra Leone or their road density. We find that, if anything, treated sections are slightly further from index cases.

We look for evidence of spillovers between treated and control sections, particularly indications that patients traveled from control to treated sections. Such reallocation within our sample would amplify our treatment effects on Ebola reporting. Assuming that patients minimize travel costs, spillovers should be largest in treated sections that border (populous) control sections. In Table E.21, we interact our treatment indicator, first, with the number of bordering control sections and, second, with the population (based on 2004 census data) in bordering control sections. If patients from control sections report in adjacent treatment areas, the coefficients on these interactions will be positive; yet, our estimates are negative and insignificant.[37] Following a similar logic, spillovers could occur via road networks. Using data from Open Street Map, we count how many control sections a treatment section is connected to via the road network (see Figure E.8). Table E.23 interacts this variable with our treatment indicator and, again, finds no indication

---

[37] We also calculate each clinic's proximity to the next nearest control clinic in the full sample. (We do not have exact coordinates of clinics, and thus geolocate clinics using the centroids of the census enumeration areas that contain the clinics.) In Table E.22 we find that treated sections report more cases when their treated clinic is far from the next control clinic — the opposite of what we would expect if spillovers are amplifying our effects.

that treated sections connected to more controls via the road network see a larger increase in total cases.

Finally, information or people may move more easily between areas that are proximate in both geographic and cultural terms. We use the household survey to determine the plurality ethnic group in each section. Sections tend to be homogeneous: in the median section, 95 percent of respondents report the same ethnicity. For each section, we then count the number of control sections with the same plurality ethnic group and within 10 kilometers. Table E.24 provides no indication that spillovers occur due to movement of patients between proximate co-ethnic areas.

4.2.2 Mechanisms

Concerns about sub-standard care deterred patients from utilizing clinics during the Ebola crisis. Fearful that seeking care would condemn their loved ones to death, households "engaged in practices of hiding sick family members, running away from local communities, or attempting to manage the course of Ebola within local households and communities" (Abramowitz et al. 2016, 24). If the CM and NFA interventions generated persistent improvements in the perceived quality of healthcare and utilization, this would help explain increased reporting in treated sections.[38] Using our endline surveys but restricting attention to the 160 clinics in the Ebola sample (see Table D.20 for estimates using our full sample), in Table 5 we estimate treatment effects on general utilization; satisfaction with public health workers; and households' beliefs about the efficacy of western ("white-man") medicine relative to traditional or religious remedies, the primary alternatives to government-run clinics in rural Sierra Leone. The effects on general utilization remain positive and significant when we pool the treatments and in CM alone; the effect attenuates in the NFA arm relative to the full sample (see Table 1). We continue to find positive effects on satisfaction, focusing here on satisfaction with public health workers, which is asked of all households.[39] Both treatment arms generate roughly equivalent increases in satisfaction with public health workers, on the order of a tenth of a standard deviation. Finally, we find sizable (0.3 standard deviation) improvements in households' attitudes towards western medicine, particularly its effectiveness relative to traditional healers or spiritual remedies. While this index is not listed among the outcomes in our analysis plan, its inclusion was motivated by assessments of the Ebola crisis stressing the importance of trust in western medicine (e.g., Kruk et al. 2015).[40]

---

[38] An alternative channel would be that improvements in physical health made people less susceptible to Ebola. However, recall that we only find health improvements for children and not adults who comprise over 70% of the confirmed Ebola cases.

[39] The satisfaction family specified in the analysis plan includes one other variable that is asked of all households at endline: whether the household is satisfied with their family's health. We do not analyze this variable, as contentment with health outcomes during "normal times" seems unlikely to shape whether one seeks care following a major adverse shock like the Ebola crisis.

[40] The importance of (dis)trust is tragically underscored by the violence facing western health providers responding to Ebola in Guinea (https://www.bbc.com/news/world-africa-29256443) and the Democratic Republic of the Congo (https://www.nytimes.com/2019/05/19/world/africa/ebola-outbreak-congo.html).

Table 5: Perceived Quality of Care

	(1) Control Mean	(2) Pooled	(3) CM	(4) NFA	(5) Difference	(6) Joint F-test (p)	(7) N
General utilization	0.000	0.066	0.122	0.025	0.096	3.819	2857
	(1.000)	(0.038)*	(0.046)***	(0.041)	(0.043)**	(0.024)**	
Satisfaction with public health workers	0.000	0.145	0.140	0.149	−0.009	4.497	3149
	(1.000)	(0.049)***	(0.065)**	(0.053)***	(0.065)	(0.013)**	
Relative effectiveness of western medicine	0.000	0.293	0.340	0.257	0.084	2.867	3183
	(1.000)	(0.123)**	(0.160)**	(0.132)*	(0.154)	(0.060)*	

Notes: Treatment effects are estimated using Missing-Indicator ANCOVA, controlling for the community-level average of the outcome family index at baseline and matching-triplet fixed effects. Robust standard errors, clustered by clinic, are shown in parentheses. Multiple-inference corrected q-values that adjust for the false discovery rate within treatment arm are shown in square brackets. Column (5) presents an F-test for the joint significance of CM and NFA, with the associated p-value in parentheses. Significance: * is significant at the 10% level, ** is significant at the 5% level and *** is significant at the 1% level.

We aggregate the three outcomes in Table 5 into a perceived quality of care index and then instrument this index with our pooled treatment indicator to estimate its effect on reported Ebola cases. In Table 6, we report a large first-stage effect; the F-statistic (9.86) approaches the rule-of-thumb for a strong instrument. When we scale our earlier reduced-form result by this first stage, we find that a one standard deviation change in the perceived quality of care corresponds to an increase in weekly case reports of 0.43.[41]

Table 6: Perceived Quality of Care and Ebola Cases

First-Stage	Perceived Quality of Care	
Pooled (CM or NFA)	0.401	
	(0.128)***	
Two-Stage Least Squares		Total Cases
Perceived Quality of Care		0.430
		(0.249)*
F-statistic	9.86	
Observations	5,440	5,440

Notes: "Perceived Quality of Care" is an equally weighted average of the variables in Table 5. Matching-triplet fixed effects included in both first- and second-stage regressions. Standard errors are clustered by section. Significance: * $p < 0.10$, ** $p < 0.05$, and *** $p < 0.01$.

As with the full sample, we do not find consistent positive effects for families focused on supply-side variables in the Ebola sample. Pooling the treatments, we see no significant effects on health service delivery or clinic quality (see Table E.14).[42]

4.3   Alternative Explanations

4.3.1   Transmission

We attribute the increase in total and confirmed cases in treated sections to reporting, not a difference in Ebola incidence. As the true incidence of Ebola in Sierra Leone is unknown — the WHO and CDC assumed they were missing half or more of all cases (Enserink 2014) — we use a number of indirect approaches to support this argument.

The interventions concluded in December 2013, five months before the first Ebola case in Sierra Leone.

---

[41] Our research design does not separately manipulate treatment status and the level of the mediator. Table 6 quantifies the effect of quality of care on reported Ebola cases only under a strong assumption about the exclusion restriction.

[42] Separating the two treatments, clinic quality increases in NFA. Unpacking the clinic quality index, the improvement does not reflect improved conditions at the clinic (e.g., cleanliness), more staff on duty, or additional hours open; rather, the increase is driven by an increase in the proportion of required services provided and a reduction in charges for out-of-stock medicine.

We find similar effects on confirmed cases from the two treatments, though NFA did not involve direct outreach to communities. The CM intervention did include community meetings. If such gatherings continued after the intervention and were sites of Ebola transmission, we would expect more infections in treatment areas to originate from contact with individuals outside of the home.[43] For a subset of infected patients, caseworkers engage in contact tracing, identifying and following up with people who may have come into contact with the infected patient. In this process, caseworkers record how these contacts relate to the patient (e.g., neighbor, tenant, brother, grandmother). In the last two columns of Table E.16, we find that patients subject to contact tracing report fewer contacts outside of their nuclear family (i.e., parents, children, siblings) in CM and NFA relative to control; the number of contacts outside of patients' nuclear families does not increase with CM.

By increasing the number of individuals reporting, the treatments could have increased contact between infected and susceptible individuals, raising the risk of nosocomial transmission.[44] To address this possibility, we compare the dates of symptom onset, reporting, and lab testing. Two features of the Ebola virus are important to note: first, Ebola incubates for 2 to 21 days (8–10 on average) before showing symptoms; and second, an individual can only test positive after displaying symptoms. Consequently, symptom onset or lab results in the first two days after a patient reports cannot not reflect infections due to exposure after the patient reports. For 92 percent of confirmed cases, symptom onset occurs prior to reporting. In 99 percent of cases (all but 2 cases), either symptom onset or laboratory testing occurs within two days of reporting, indicating that nearly all confirmed cases we count do not result from infections that occur after the case was reported.[45]

As further evidence against nosocomial transmission in our sample, Fang et al. (2016) report that infections among healthcare workers fall precipitously by September 2014 (the start of our study period), indicating improved awareness and infection control. We continue to find treatment effects in the months after a nationwide effort to train healthcare workers in isolation and no-touch treatment (see Appendix Table E.3).

Finally, we look at the ratio of confirmed to total cases across treatment and control areas to determine whether the interventions increased the share of infected patients among total cases. This ratio is however undefined when no cases are reported in a section-week. Below, we take a bounding approach, imputing

---

[43] Community gatherings are unlikely sites of Ebola contagion. The virus is transmitted through direct contact with infected bodily fluids (blood, feces, semen, spit, sweat, vomit). Funerals were sites of transmission, because participants touched infected corpses. Unlike airborne pathogens, proximity is not sufficient to spread Ebola: Glynn et al. (2018) estimate a secondary attack rate of only 18 percent among individuals living in the same household as a confirmed Ebola patient.

[44] See also Lowes and Montero (2018) who demonstrate long-lasting unintended consequences from nosocomial transmission in the context of historical public health campaigns in West Africa, which highlight the importance of addressing this alternative account.

[45] The proportions are nearly identical among patients who test negative for Ebola: 89.8 percent have symptom onset prior to reporting, and 99.4 percent have onset or lab testing within 2 days of reporting.

either all ones or all zeros to observations where the ratio is undefined. Imputing all ones assumes that, if cases had reported, they would have all been confirmed; imputing all zeros assumes that, if cases had reported, none would have tested positive. Figure E.9(a) plots the average ratio of confirmed to total cases across control and treated sections. Looking at either bound, there is no meaningful difference in these ratios, and the confidence intervals overlap throughout the study period.[46]

In Section E.20, we write down a model to clarify what must be assumed for our results to reflect a change in exposure (as opposed to reporting). For confirmed cases to increase while the share of confirmed to total to remains unchanged, one must conjecture that the treatments dramatically increased reporting by asymptomatic individuals while having negligible effects among those showing possible signs of the virus. This strains credulity: one cannot preemptively test for Ebola, so individuals without symptoms have no reason to report; moreover, qualitative accounts suggest the crisis deterred unexposed individuals from visiting clinics, even when they had other healthcare needs (Elston et al. 2016).

4.3.2 Surveillance

Surveillance is challenging in a setting like rural Sierra Leone: if individuals or families want to avoid health workers, they are unlikely to be detected given the difficulties of canvassing sparsely populated regions with limited road networks (Richards et al. 2015; Olu et al. 2016; McNamara et al. 2016). Nonetheless, we look for indications of intensified top-down efforts in treated sections as a possible alternative explanation for the increase in reported cases.

First, contact tracing is central to disease surveillance efforts. In our control sections 59 percent of confirmed cases were subject to contact tracing, compared to just 22 percent in CM and 28 percent in NFA (Table E.16). Second, we use three measures derived from the VHF data (all measured at the section-level): (1) the probability that a case received laboratory testing to confirm or rule out an infection; (2) the average number of days that passed between a case being reported and lab testing; and (3) the number of unique case workers (logged) that entered information into the VHF. We expect these variables to proxy for top-down surveillance efforts during the crisis. In Table E.17, we find no significant differences for these variables across treatment and control.

Finally, using data from Sierra Leone's National Ebola Response Center (NERC) and the UN Mission for Ebola Emergency Response (UNMEER), we count the number of Ebola-specific treatment facilities in each section (see Section E.1). There were three types of specialized facilities: Ebola Treatment Units

---

[46] It is possible that the ratio of confirmed to total cases could stay constant if there was an increase in the number of probable and suspected cases. Figure E.9(b) repeats the bounding exercise but uses the ratio of confirmed to confirmed plus negative cases. This exercise delivers the same conclusion, as the number of probable and suspected cases are small and unaffected by treatment (see Table E.6.

(ETU, 32 beds on average), Ebola Holding Centers (EHC, 18 beds on average), and Community Care Centers (CCC, 10 beds on average). Only one ETU falls within our sample, and it is located in a control section; Table E.18 shows no significant difference in the counts — either combined or separate — of EHCs or CCCs.[47]

Our results in Table 3 are robust to dropping the small number of sections that contain one or more of these specialized facilities: when analyzing total reported cases, the coefficient on our pooled treatment indicator increases from 0.173 to 0.177 (se = 0.091) when we drop these sections; for confirmed cases, it changes from 0.59 to 0.55 (se = 0.025).[48]

5. Conclusion

This study shows that accountability interventions, such as community monitoring and non-financial awards, can boost the perceived quality of health care and improve health outcomes in a developing country setting — not only during "normal" times, but also during crises. We use a randomized experiment completed less than one year before the Ebola outbreak in Sierra Leone to test the effectiveness of two interventions: one focused on community monitoring of government-run health clinics and the other on status awards for clinic staff. In the medium term, we find that both interventions improve utilization of clinics, patient satisfaction, and beliefs about the relative efficacy of western medicine. Similar to previous work, we find large decreases in under-5 mortality following the community monitoring intervention (Björkman and Svensson 2009). During the Ebola crisis, we find that both interventions substantially increased reporting — by both patients who test positive and negative for the virus — and reduced mortality among patients.

Thus, we find evidence that accountability interventions which leverage social incentives to increase the perceived quality of health care can lead to greater utilization of health systems in a low-resource setting, in ways that help build their resiliency to confront crises.

We explore two alternative explanations for why the interventions increased reporting during the Ebola crisis: by unintentionally increasing exposure to Ebola; or by enabling more top-down surveillance efforts. We do not find support for these mechanisms. Specifically, we find no evidence to indicate that the interventions contributed to transmission at treated clinics, or that they raised the infection rate among patients

---

[47] The absolute numbers here are instructive: there is 1 EHC in control sections, 1 in NFA sections, and 2 in CM sections. In Figures E.4 and E.5 we drop all triplets and pairs of triplets as a robustness check to address concerns that a small number of sections could drive our results.

[48] The locating of specialized facilities in nearby sections could depress reported cases, as patients might report directly to those facilities and, thus, not be counted within their home section. In Table E.19 we find that treated sections are not significantly further from ETUs, EHCs or CCCs in the NERC data; the distance from NFA sections to the nearest CCCs is shorter when we use the UNMEER data.

(i.e., the ratio of confirmed to total cases). Rather, the interventions increased reporting by both infected individuals, as well as those who feared they had been exposed but tested negative. We also see no indication of more Ebola-specific treatment facilities, lab resources, or caseworkers in treated areas, suggesting that resources for screening and contact-tracing were not targeted to areas that received the interventions.

For policy-makers, our findings suggest that in low-resource contexts, accountability interventions can be leveraged to improve utilization, satisfaction, and patient health. Our estimates for longer-run, Ebola-related outcomes show that the impact of such interventions can extend to — and may even be amplified in — crisis settings. To the best of our knowledge, we present the first experimental evidence demonstrating how accountability interventions affect crises outcomes. Our results suggest that by encouraging patients to report and seek treatment, which back-of-the-envelope calculations suggest reduced the disease's reproduction rate by 19 percent (see Appendix Section E.23), these simple interventions can improve a health system's capacity to weather crises, a marker of resiliency. These results align with other recent work suggesting that social factors, including low levels of trust in the health system, are of crucial importance in crisis settings such as the Ebola outbreak, where individuals face a choice about whether or not to cooperate with containment efforts (Blair et al. 2017; Morse et al. 2016; Tsai et al. 2019; Vinck et al. 2019). In fact, decisions around whether to cooperate also apply to a wide range of epidemic contexts, including the global outbreak of the novel coronavirus (COVID-19). As such, our findings, which highlight how boosting public confidence can help contain crises, hold potential implications for a broader range of crises beyond Ebola.

Social interventions like the ones we study here could complement more traditional crisis-response efforts. For instance, during the 2014–15 Ebola crisis, the WHO designed and deployed eight-bed clinics called Community Care Centers (CCC). The purpose of these centers was to allay fears about western medical facilities and, thus, encourage reporting and early isolation and treatment (Michaels-Strasser et al. 2015). In a quasi-experimental evaluation of CCCs, we find that they are successful in encouraging reporting and isolation during the crisis; in fact, they have effects on reports of confirmed cases that are larger, though of the same order of magnitude, as the interventions we study here (Christensen et al. forthcoming). Future work might examine the relative cost-effectiveness of the two types of interventions, as well as potential complementarities.

References

Abramowitz, Sharon, Braeden Rogers, Liya Akilu, Sylvia Lee, and David Hipgrave (2016, March). "Ebola Community Care Centers: Lessons learned from UNICEF's 2014-2015 Experience in Sierra Leone." *Maternal, Newborn, and Child Health Working Paper*.

Alsan, Marcella and Marianne Wanamaker (2017, 08). "Tuskegee and the Health of Black Men." *The Quarterly Journal of Economics* 133(1), 407–455.

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.

Andrabi, Tahir, Jishnu Das, Asim I Khwaja, Selcuk Ozyurt, and Niharika Singh (2018). *Upping the ante: The equilibrium effects of unconditional grants to private schools*. The World Bank.

Ashraf, Nava and Oriana Bandiera (2018). "Social incentives in organizations." *Annual Review of Economics* 10, 439–463.

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.

Bandiera, Oriana, Niklas Buehren, Markus Goldstein, Imran Rasul, and Andrea Smurra (2019). "The economic lives of young women in the time of ebola." *World Bank Policy Research Working Paper*.

Banerjee, Abhijit, Angus Deaton, and Esther Duflo (2004). "Health care delivery in rural rajasthan." *Economic and Political Weekly*, 944–949.

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

Banerjee, Abhijit V, Esther Duflo, and Rachel Glennerster (2008). "Putting a band-aid on a corpse: incentives for nurses in the Indian public health care system." *Journal of the European Economic Association* 6(2-3), 487–500.

Barr, Abigail, Frederick Mugisha, Pieter Serneels, and Andrew Zeitlin (2012). "Information and collective action in community-based monitoring of schools: Field and lab experimental evidence from Uganda." *Unpublished paper, Georgetown University*.

Benabou, Roland and Jean Tirole (2003). "Intrinsic and extrinsic motivation." *The Review of Economic Studies* 70(3), 489–520.

Besley, Timothy and Maitreesh Ghatak (2005). "Competition and incentives with motivated agents." *American Economic Review* 95(3), 616–636.

Björkman, Martina and Jakob Svensson (2009, May). "Power to the People: Evidence from a Randomized Field Experiment on Community-Based Monitoring in Uganda." *Quarterly Journal of Economics* 124(2), 735–769.

Björkman Nyqvist, Martina, Damien de Walque, and Jakob Svensson (2017, January). "Experimental Evidence on the Long-Run Impact of Community-Based Monitoring." *American Economic Journal: Applied Economics* 9(1), 33–69.

Blair, Robert A., Benjamin S. Morse, and Lily L. Tsai (2017). "Public health and public trust: Survey evidence from the ebola virus disease epidemic in liberia." *Social Science & Medicine* 172, 89 – 97.

Bostock, Bill (2020). "A city in hubei, china, is giving residents $1,400 if they report their coronavirus symptoms to doctors and then test positive." *Business Insider* (https://www.businessinsider.com/coronavirus-china-hubei-qianjiang-city-reward-reporting-symptoms-test-positive-2020-2).

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.

CDC (2019). "Cost of the ebola epidemic." (https://www.cdc.gov/vhf/ebola/history/2014-2016-outbreak/cost-of-ebola.html).

Christensen, Darin, Dube Oeindrila, Johannes Haushofer, Bilal Siddiqi, and Maarten Voors (forthcoming). "Community-based Crisis Response: Evidence from Sierra Leone's Ebola Outbreak." *AEA Papers and Proceedings*.

Das, Jishnu, Alaka Holla, Aakash Mohpal, and Karthik Muralidharan (2016). "Quality and accountability in health care delivery: Audit-study evidence from primary care in india." *American Economic Review* 106(12), 3765–99.

Denney, Lisa and Richard Mallett (2014, September). "Mapping Sierra Leone's plural health system and how people navigate it." Technical report.

Dixit, Avinash et al. (2002). "Incentives and organizations in the public sector: An interpretative review." *Journal of Human Resources* 37(4), 696–727.

Dupas, Pascaline (2011). "Health behavior in developing countries." *Annual Review of Economics* 3(1), 425–449.

Dupas, Pascaline and Edward Miguel (2017). "Impacts and determinants of health levels in low-income countries." In *Handbook of economic field experiments*, Volume 2, pp. 3–93. Elsevier.

Elston, J W T, A J Moosa, F Moses, G Walker, N Dotta, R J Waldman, and J Wright (2016, December). "Impact of the Ebola outbreak on health systems and population health in Sierra Leone." *Journal of Public Health* 38(4), 673–678.

Enserink, Martin (2014). "How many Ebola cases are there really?" https://www.sciencemag.org/news/2014/10/how-many-ebola-cases-are-there-really, Accessed in September 2019.

Fang, Li-Qun, Yang Yang, Jia-Fu Jiang, Hong-Wu Yao, David Kargbo, Xin-Lou Li, Bao-Gui Jiang, Brima Kargbo, Yi-Gang Tong, Ya-Wei Wang, Kun Liu, Abdul Kamara, Foday Dafae, Alex Kanu, Rui-Ruo Jiang, Ye Sun, Ruo-Xi Sun, Wan-Jun Chen, Mai-Juan Ma, Natalie E Dean, Harold Thomas, Ira M Longini Jr, M Elizabeth Halloran, and Wu-Chun Cao (2016, April). "Transmission dynamics of Ebola virus disease and intervention effectiveness in Sierra Leone." *Proceedings of the National Academy of Sciences* 113(16), 4488–4493.

Fiala, Nathan and Patrick Premand (2018). "Social accountability and service delivery: Experimental evidence from uganda." *World Bank Policy Research Working Paper* (8449).

Finan, Frederico, Benjamin A. Olken, and Rohini Pande (2017). "The personnel economics of the developing state." *Handbook of Field Experiments* II, 467–514.

Garske, Tini, Anne Cori, Archchun Ariyarajah, Isobel M Blake, Ilaria Dorigatti, Tim Eckmanns, Christophe Fraser, Wes Hinsley, Thibaut Jombart, Harriet L Mills, Gemma Nedjati-Gilani, Emily Newton, Pierre Nouvellet, Devin Perkins, Steven Riley, Dirk Schumacher, Anita Shah, Maria D Van Kerkhove, Christopher Dye, Neil M Ferguson, and Christl A Donnelly (2017, April). "Heterogeneities in the case fatality ratio in the West African Ebola outbreak 2013–2016." *Philosophical Transactions of the Royal Society B: Biological Sciences* 372(1721), 20160308.

Glynn, Judith R, Hilary Bower, Sembia Johnson, Cecilia Turay, Daniel Sesay, Saidu H Mansaray, Osman Kamara, Alie Joshua Kamara, Mohammed S Bangura, and Francesco Checchi (2018, January). "Variability in Intrahousehold Transmission of Ebola Virus, and Estimation of the Household Secondary Attack Rate." *The Journal of Infectious Diseases* 217(2), 232–237.

Karing, Anne (2019). *Social Signaling and Health Behavior in Low-Income Countries*. Ph.D. thesis, UC Berkeley.

Kling, Jeffrey R, Jeffrey B Liebman, et al. (2004). "Experimental analysis of neighborhood effects on youth." *Working paper*.

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

Kosfeld, Michael and Susanne Neckermann (2011, August). "Getting More Work for Nothing? Symbolic Awards and Worker Performance." *American Economic Journal: Microeconomics* 3(3), 86–99.

Kruk, Margaret E, Anna D Gage, Catherine Arsenault, Keely Jordan, Hannah H Leslie, Sanam Roder-DeWan, Olusoji Adeyi, Pierre Barker, Bernadette Daelmans, Svetlana V Doubova, et al. (2018). "High-quality health systems in the sustainable development goals era: Time for a revolution." *The Lancet Global Health* 6(11), e1196–e1252.

Kruk, Margaret E, Michael Myers, S Tornorlah Varpilah, and Bernice T Dahn (2015, May). "What is a resilient health system? Lessons from Ebola." *The Lancet* 385(9980), 1910–1912.

Levy, Benjamin, Carol Y Rao, Laura Miller, Ngozi Kennedy, Monica Adams, Rosemary Davis, Laura Hastings, Augustin Kabano, Sarah D Bennett, and Momodu Sesay (2015, July). "Ebola infection control in Sierra Leonean health clinics: A large cross-agency cooperative project." *American Journal of Infection Control* 43(7), 752–755.

Lowes, Sara and Eduardo Montero (2018). "The legacy of colonial medicine in central africa." *Working Paper* (https://scholar.harvard.edu/slowes/publications/colonial-medicine).

Mansuri, Ghazala and Vijayendra Rao (2003). "Localizing development: Does participation work?" Technical report, World Bank.

Markham, Steven E, K Dow Scott, and Gail H McKee (2002). "Recognizing good attendance: A longitudinal, quasi-experimental field study." *Personnel Psychology* 55(3), 639–660.

McNamara, Lucy A, Ilana J Schafter, Leisha D Nolen, Yelena Gorina, John T Redd, Terrence Lo, Elizabeth Ervin, Olga Henao, Benjamin A Dahl, Oliver Morgan, Sara Hersey, and Barbara Knust (2016). "Ebola Surveillance—Guinea, Liberia, and Sierra Leone." *MMWR supplements* 65(3), 35–43.

Michaels-Strasser, Susan, Miriam Rabkin, Maria Lahuerta, Katherine Harripersaud, Roberta Sutton, Laurence Natacha Ahoua, Bibole Ngalamulume, Julie Franks, and Wafaa M El-Sadr (2015, July). "Innovation to confront Ebola in Sierra Leone: The community-care-centre model." *The Lancet Global Health* 3(7), e361–e362.

Miller, Grant, Renfu Luo, Linxiu Zhang, Sean Sylvia, Yaojiang Shi, Patricia Foo, Qiran Zhao, Reynaldo Martorell, Alexis Medina, and Scott Rozelle (2012). "Effectiveness of provider incentives for anaemia reduction in rural china: A cluster randomised trial." *Bmj* 345, e4809.

Morse, Ben, Karen A Grépin, Robert A Blair, and Lily Tsai (2016). "Patterns of demand for non-ebola health services during and after the ebola outbreak: Panel survey evidence from monrovia, liberia." *BMJ Global Health* 1(1), e000007.

Mozur, Paul (2020). "China, desperate to stop coronavirus, turns neighbor against neighbor." *New York Times* (https://www.nytimes.com/2020/02/03/business/china-coronavirus-wuhan-surveillance.html).

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

Olken, Benjamin A, Junko Onishi, and Susan Wong (2014). "Should aid reward performance? evidence from a field experiment on health and education in indonesia." *American Economic Journal: Applied Economics* 6(4), 1–34.

Olu, Olushayo Oluseun, Margaret Lamunu, Miriam Nanyunja, Foday Dafae, Thomas Samba, Noah Sempiira, Fredson Kuti-George, Fikru Zeleke Abebe, Benjamin Sensasi, Alexander Chimbaru, Louisa Ganda, Khoti Gausi, Sonia Gilroy, and James Mugume (2016). "Contact tracing during an outbreak of ebola virus disease in the western area districts of sierra leone: Lessons for future ebola outbreak response." *Frontiers in Public Health* 4, 130–9.

Owada, Kei, Tim Eckmanns, Kande-Bure O'Bai Kamara, and Olushayo Oluseun Olu (2016, August). "Epidemiological Data Management during an Outbreak of Ebola Virus Disease: Key Issues and Observations from Sierra Leone." *Frontiers in Public Health* 4(46), 1064–4.

Pradhan, Menno, Daniel Suryadarma, Amanda Beatty, Maisy Wong, Armida Alishjabana, Arya Gaduh, and Rima Prama Artha (2011). *Improving educational quality through enhancing community participation: Results from a randomized field experiment in Indonesia*. The World Bank.

Pronyk, Paul, Braeden Rogers, Sylvia Lee, Aarunima Bhatnagar, Yaron Wolman, Roeland Monasch, David Hipgrave, Peter Salama, Adam Kucharski, Mickey Chopra, and on behalf of the UNICEF Sierra Leone Ebola Response Team (2016, April). "The Effect of Community-Based Prevention and Care on Ebola Transmission in Sierra Leone." *American Journal of Public Health* 106(4), 727–732.

Raffler, Pia, Daniel N Posner, and Doug Parkerson (2019). "The weakness of bottom-up accountability: Experimental evidence from the ugandan health sector." *Working Paper*.

Raven, Joanna, Haja Wurie, and Sophie Witter (2018). "Health workers' experiences of coping with the ebola epidemic in sierra leone's health system: A qualitative study." *BMC Health Services Research* 18(1), 251–260.

Richards, Paul, Joseph Amara, Mariane C Ferme, Prince Kamara, Esther Mokuwa, Amara Idara Sheriff, Roland Suluku, and Maarten Voors (2015, April). "Social Pathways for Ebola Virus Disease in Rural Sierra Leone, and Some Implications for Containment." *PLOS Neglected Tropical Diseases* 9(4), 1–15.

Singh, Prakarsh and Sandip Mitra (2017). "Incentives, information and malnutrition: Evidence from an experiment in India." *European Economic Review* 93, 24–46.

The World Bank (2003). *World Development Report 2004: Making Services Work for Poor People.* World Bank.

Tsai, Lily, Benjamin Morse, and Robert Blair (2019). "Building trust and cooperation in weak states: Persuasion and source accountability in liberia during the 2014-2015 ebola crisis." *Working paper*.

UNICEF (2014, December). "Sierra Leone Health Facility Survey 2014: Assessing the impact of the EVD outbreak on health systems in Sierra Leone." Technical report, UNICEF.

Vandi, M A, J van Griensven, A K Chan, B Kargbo, J N Kandeh, K S Alpha, A A Sheriff, K S B Momoh, A Gamanga, R Najjemba, and S Mishra (2017, June). "Ebola and community health worker services in Kenema District, Sierra Leone: Please mind the gap!" *Public Health Action* 7(Suppl 1), S55–61.

Vinck, Patrick, Phuong N Pham, Kenedy K Bindu, Juliet Bedford, and Eric J Nilles (2019). "Institutional trust and misinformation in the response to the 2018-19 ebola outbreak in north kivu, dr congo: A population-based survey." *The Lancet Infectious Diseases* 19, 529–536.

Waxman, Matthew, Adam R Aluisio, Soham Rege, and Adam C Levine (2017, June). "Characteristics and survival of patients with Ebola virus infection, malaria, or both in Sierra Leone: A retrospective cohort study." *The Lancet Infectious Diseases* 17(6), 654–660.

Wen, Leana S (2020). "Governments need people's trust to stop an outbreak. where does that leave us?" *Washington Post* (https://www.washingtonpost.com/opinions/2020/01/22/governments-need-peoples-trust-stop-an-outbreak-where-does-that-leave-us/).

Whitty, Christopher J M, Jeremy Farrar, Neil Ferguson, W John Edmunds, Peter Piot, Melissa Leach, and Sally C Davies (2014, November). "Infectious disease: Tough choices to reduce Ebola transmission." *Nature* 515(7526), 192–194.

WHO (2014). "Experimental therapies: growing interest in the use of whole blood or plasma from recovered ebola patients (convalescent therapies)." (https://www.who.int/mediacentre/news/ebola/26-september-2014/en/).

Supporting Information

BUILDING RESILIENT HEALTH SYSTEMS:
EXPERIMENTAL EVIDENCE FROM SIERRA LEONE AND THE 2014 EBOLA OUTBREAK

Following text to be published online.

Contents

A. Context

A.1 Administrative Boundaries

Figure A.1: Administrative Boundaries

(a) Districts

(b) Chiefdoms
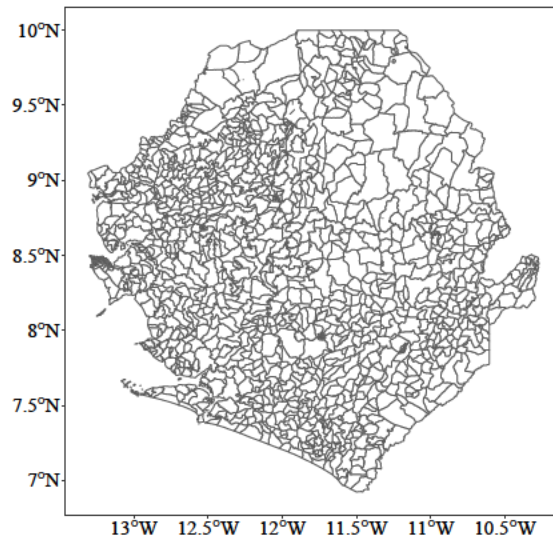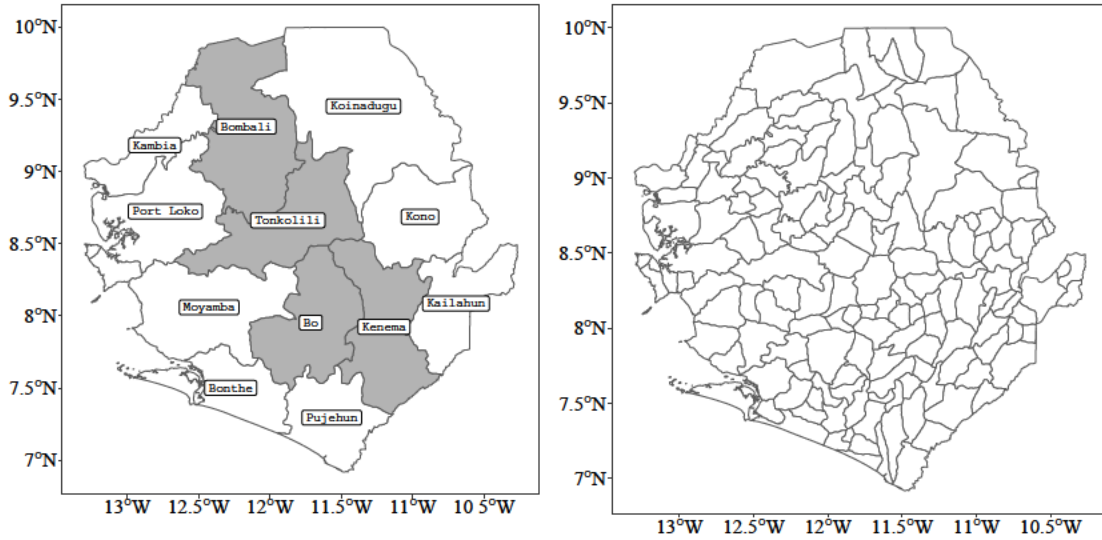
(c) Sections

Maps of Sierra Leone's different administrative units are provided for reference. Sections (bottom) nest neatly in chiefdoms (top right), which nest neatly in districts (top left). The original randomized experiment was run in four districts: Bombali, Tonkolili, Bo, and Kenema.

A.2 Cross-National Health Indicators

Figure A.2: Health Expenditure and Under-5 Mortality in 2010
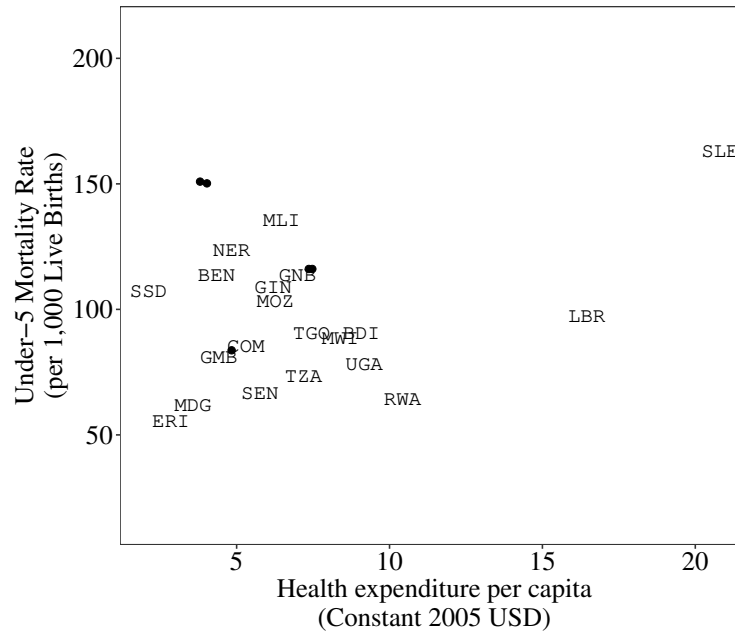
We use measures from the World Development Indicators from 2010 for under-5 mortality (per 1,000 live births) and health expenditure per capita (in constant 2005 USD). The sample includes countries that the World Bank classifies as low income. Sierra Leone (SLE) appears to the upper right.

B. Variable descriptions

B.1 Pre-specified outcome variables and deviations

Below we detail the families of outcome variables analyzed in the medium-term results. Outcome variables are marked by {i} if measured at the individual level, by {hh} if measured at the household-level, by {com} if measured at the community/village-level and by {phu} if measured at the clinic-level. Where outcomes analyzed deviate from our pre-specified analysis plan, we detail the deviation using footnotes. In accounting for multiple comparisons, we deviate from the original analysis plan in the conservative direction. Originally, we only planned to control for the false discovery rate (FDR) across just a subset of families, and within each family, a subset of indicators. We instead make the FDR adjustment across all families, and within each family, across all indicators.

1. **General Utilization Index**

    (a) Health episodes in response to which individuals visited clinic {i} [A1]

        i. Proportion of all health episodes in response to which individual visited traditional healer (among individuals experiencing health episodes)

        ii. Which health providers did you visit? (among individuals experiencing health episodes)

2. **Maternal Utilization Index (maternal episodes)**

    (a) Antenatal/postnatal care index [standardized summary index of i-ii] {i}

        i. Number of ANC visits (among mothers who have given birth in the last year)

        ii. Number of PNC visits (among mothers who have given birth in the last year)

    (b) Childbirth in facility {i}

        i. Proportion of pregnant mothers who gave birth in facility (among mothers who have given birth in the last year)

3. **Health Outcomes**

    (a) Proportion of households where at least one child under the age of 5 has died (in the past 6 months) {hh}

[A1]The original analysis plan defined general utilization as an index composed of both the utilization of western medical clinics (entering the index positively) and utilization of traditional or religious healers (entering negatively). We later found that the survey questions from which we intended to obtain information on use of traditional or religious healers was unsuitable for this purpose. In particular, only the illness / injury module asked utilization questions which explicitly included the traditional healers / religious or spiritual leaders as an option category. For the other three types of health episodes (child birth, vaccinations and ANC/PNC visits), the answer options contained a health provider category of "other", which could not be (unambiguously) attributed to traditional healers. We therefore restrict our utilization variable to utilization of western medical clinics.

(b) Proportion of households where women have died during OR due to complications from pregnancy (in the past 6 months) {hh}

(c) Proportion of households where any household member had an illness {hh}

    i. Was this episode an illness, an injury or other consultation?

(d) Anthropometric outcomes {hh}

    i. Child weight-for-height (Among eligible children. Measured at endline only)

(e) Vaccine completion index: (Among households with eligible children) [standardize d summary index of A - G] {hh}

    i. Proportion of children in household completing full cycle of: (A) BCG, (B) OPV, (C) Penta, (D) Measles, (E) Yellow Fever, (F) RVV, (G) PCV

(f) Childbirth episode [standardized summary index of i - ii] {hh}

    i. Did the mother have health problems during or within two months of the delivery?

    ii. Did the baby have health problems during delivery or within one month of birth?

(g) Child illness index [standardized summary index of i - ii] {phu}

    i. Number of malaria cases (among children under 5)

    ii. Number of diarrhea cases (among children under 5)

4. Satisfaction

(a) How satisfied are you with your family's health? {hh}

(b) How satisfied are you with the performance of public health workers? {hh}

(c) Satisfaction with services {hh}

    i. The last time you visited [CLINIC] in the past one month, how satisfied were you with the care that you received at the clinic?

    ii. The next time you need medical attention for some other reason, would you visit [CLINIC] again?

5. Clinic Quality

(a) Clinic service provision [standardized summary index of i - vi] {phu}

    i. Facility organization index [standardized summary index of A-R]

        A. (A) Duty Roster for Staff, (B) Numbered cards for patients, (C) Seating Arrangements, (D) Suggestion box, (E) Name tags for staff, (F) Rooms labeled, (G) Floor clean, (H) Walls clean, (I) Area clean/uncluttered, (J) Drug info available, (K) Smells okay, (L) Coverage graphs, (M) Medicines on floor, (N) Medicines organized by date, (O) Drugs stored in safe area, (P) Storage room clean, (Q) Storage room has limited access, (R) Stock cards available

ii. Proportion of required services provided by clinic (In the past month) [proportion of A-L the clinic is required to provide]

   A. (A) Immunization, (B) Growth monitoring, (C) Treatment of sick children, (D) Antenatal care, (E) Family planning, (F) Treatment of STIs/STDs, (G) Deliveries (enumerator ask anything associated with delivery e.g. soap, incentive for TBAs), (H) HIV / AIDS counseling and testing (I) Health education, (J) Postnatal care, (K) Nutrition supplementation, (L) Pregnancy test

iii. Frequency of service provision index [standardized summary of the number of days (ii) are provided]

iv. Proportion of clinics charging for out of stock equipment

v. Number of clinic workers on duty

vi. Reported hours clinic is open (per week)

(b) Proportion of clinics that know about the free health-care policy {phu}

(c) Employee satisfaction index [standardized summary index of i-ii] {phu}

   i. Satisfaction with community support/participation

   ii. Satisfaction with job overall

6. Health Service Delivery

(a) Absenteeism (among respondents experiencing health episodes) [standardized summary index of i-ii] {i}

   i. Of all the times that you visited the clinic in the past one month, did you ever find there was no staff present?

   ii. The last time you visited the clinic in the past one month, how long did you wait to see the person who attended to you?

(b) Fee payments (among all health episodes) {i}

   i. Did you pay any money for products or services during this consultation?

   ii. What is the total estimated value of the items (in cash and in kind) that you gave the person/people who assisted you?

(c) Service delivery (among all health episodes) {i}

   i. In the past one month, have you had any problems with the clinic?

   ii. What were these problems?

      A. Staff not present

      B. Drugs not available

      C. Facility not clean

      D. Unpleasant behaviour from staff

(d) Were medicines in-stock and available at the clinic? (among all health episodes) {i}

(e) Satisfaction with services {i}

    i. The last time you visited the clinic in the past one month, how satisfied were you with the care that you received at the clinic?

    ii. The next time you need medical attention for some other reason, would you visit [CLINIC] again?

(f) The last time you visited the clinic in the past one month, how would you rate the attitude of the staff? {i}

7. **Community Support**

(a) Reported engagement index [standardized summary index of i-iii] {com}

    i. Health monitoring facility (HMF)/clinic monitoring facility (CMF) exists

    ii. Number of HMC/FMC meetings

    iii. Contributions to clinic (e.g. expenditures, nurse veg garden, etc.)[A2]

(b) Reported community engagement index (past 6 months)[standardized summary index of ii-vi] {phu}

    i. Has the community helped clean this facility?

    ii. Has the community helped you with your personal work? E.g. Farm, back garden... etc.

    iii. How often have community members helped you with your personal work?

    iv. How often has the facility had disputes/conflicts with the community?

8. **Community Development and Political Engagement (CDPE)[A3]**

(a) Development projects (Excluding NGOs) {com}[A4]

    i. Has [the Local Council/the Paramount Chief] done any projects that this community (In the past year, starting May 2012)

    ii. Did community members contribute labour, money or local materials for this project (Including work for food and work for pay)?

    iii. Were any community members involved in the planning of this project?

(b) Collective action {com}

    i. Has this community worked together to address any problem facing this community? For each project: (In the past one year since May 2012)

[A2] The analysis plan mis-specified financial contributions as originating from the community survey, but it was part of the clinic survey and is included in index ii accordingly.

[A3] Originally, both the CDPE and Community Support indices included the HMC / HMF meetings variables. We retain these as a part of the Community Support index, as this index is intended to more directly gauge the monitoring mechanism more directly. However, we verify that omitting these variables from the CDPE index has no consequence on the estimated effect. (These results are available upon request).

[A4] The baseline survey asked for projects in the past two years while the endline survey asked for the past year. In the ANCOVA specification, we therefore control for projects in the past two years.

A. What kind of problem did this community address?

B. Did the community approach any person or organization outside the community for help in addressing this problem?

C. Whom did the community first approach regarding this problem?

D. Is your community satisfied with the way in which the person / organization responded to your problem?

E. Has this problem now been resolved?

(c) Voting {hh}

    i. Do you have a voter registration card?

    ii. Did you vote in the last Local Council Elections? (November 2012 election)

    iii. Did you vote in the last General Elections? (November 2012 election)

9. Water and Sanitation

(a) Household-level index [standardized summary index of i-ii] {hh}

    i. Water

        A. What is the main source of drinking water for members of your household?

        B. What do you usually do to make the water safer to drink?

        C. What is the main source of water used by your household for other purposes such as cooking and hand washing?

    ii. Toilets

        A. What type of toilet facility do members of your household usually use?

(b) Community-level index [standardized summary index of i-ii] {com}

    i. Water

        A. Is there a water facility in this village/community?

        B. What kind of water facility is it?

        C. Do people from this community usually get water to drink from this water facility?

        D. [If not] Where do people from this community usually get water to drink?

    ii. Toilets

        A. Is there a Communal Waste Disposal site in this village?

        B. Are there any public toilets in your community?

(c) Satisfaction index {hh}

    i. How satisfied are you with the public health and sanitation facilities such as drainage, toilets, garbage bins and access to clean and safe water?

    ii. How satisfied are you with the cleanliness of your community?

    iii. Over the last year how has the quality of public health and sanitation changed?

10. **Economic Status**

   (a) Physical asset index: {hh}

      i. How many of the following does this household own in either usable or repairable condition? a) Generator, b) Radio, c) Television, d) Mobile, Telephone, e) Non-mobile Telephone, f) Refrigerator, g) Electric Fan, h) Watch or Clock, i) Umbrella, j) Large Cooking Pot, k) A Bicycle, l) A Motorcycle or Motor scooter, m) An animal-drawn cart, n) A Car or Truck, o) A Boat with no Motor, p) A Boat with a Motor

   (b) Agricultural asset index {hh}

      i. At present, how many agricultural assets does this household own in either usable or repairable condition? E.g. hoe, cutlass, shovel, spade, sickle, plough, cassava grater, thresher etc. (Dichotomized based on presence)

      ii. For each of the animals below, ask "How many "_____" do members of the household own?" a) Cows/Bulls, b) Horses/Donkeys, c) Pigs, d) Goats, e) Sheep, f) Rabbits, g) Rodents, h) Fowl (Chickens), i) Ducks, j) Other Birds. (Dichotomized based on presence)

   (c) Dwelling materials index {hh}

      i. What is the main material of the floors of the house?

      ii. What is the material of the roof of the house?

      iii. What is the material of the exterior walls of the house?

   (d) Total consumption expenditure {hh}

      i. How much in total have members of your household spent on "_____" (In the past month)?

      ii. Did you consume "_____" from your own harvest or your own stock in the past month?

         A. How much of "_____" did you consume in the past month?[A5]

[A5] In the original analysis plan we also intended to include prices, based on the question, "For how much would you sell this amount of "_____" if you were to sell it now?" However, these prices were not properly recorded and are therefore omitted from the index.

B.2 Descriptive statistics

Table B.1: General Utilization

	Median	Mean	SD	Min	Max	N	BL data
(1) Number of health episodes in which sought western care ① ⑫ ◇	1	0.988	0.382	0	3	4496	Yes

Questions about ante and post natal care, vaccinations and illness/injury episodes are asked for the past one month. Questions about child birth are asked about the past year. †: dummy variable; §: likert scale (1-4); ¶: likert scale (1-5); ◇: measured across all health episode types (ante and post natal care, child birth, under-2 vaccinations, and illness and injury); ⓦ, ①, ⑥, and ⑫ stand for measures referring to the last week, one month, six months and twelve months respectively.

Table B.2: Maternal Utilization

	Median	Mean	SD	Min	Max	N	BL data
(1) ANC/PNC visits index ⑫	−0.187	−0.054	0.956	−1.779	5.602	887	Yes
(2) Birth in western medicine facility ⑫ †	1	0.862	0.345	0	1	877	Yes

†: dummy variable; §: likert scale (1-4); ¶: likert scale (1-5); ◇: measured across all health episode types (ante and post natal care, child birth, under-2 vaccinations, and illness and injury); ⓦ, ①, ⑥, and ⑫ stand for measures referring to the last week, one month, six months and twelve months respectively.

Table B.3: Health Outcomes

	Median	Mean	SD	Min	Max	N	BL data
(1) U5 Child death per HH ⑥ †	0	0.033	0.178	0	1	5053	Yes
(2) Maternal death per HH ⑥ †	0	0.001	0.034	0	1	5053	Yes
(3) Illness/injury in HH ① †	1	0.583	0.493	0	1	5053	Yes
(4) Child weight for length	0.600	0.621	1.633	−4.910	4.980	1991	No
(5) Vaccine completion index (Under 2)	4	3.117	2.589	0	7	1457	Yes
(6) Child birth complication index ⑫	−0.619	−0.002	0.975	−0.619	2.649	856	Yes
(7) Child illness index ①	−0.091	0.028	0.924	−1.435	3.424	4993	Yes

Maternal death is defined as death relating to either pregnancy complications or childbirth. The Pentavalent Vaccination targets diphtheria, tetanus, whooping cough, hepatitis B, and haemophilus influenza type B. †: dummy variable; §: likert scale (1-4); ¶: likert scale (1-5); ◇: measured across all health episode types (ante and post natal care, child birth, under-2 vaccinations, and illness and injury); ⓦ, ①, ⑥, and ⑫ stand for measures referring to the last week, one month, six months and twelve months respectively.

Table B.4: Satisfaction

	Median	Mean	SD	Min	Max	N	BL data
(1) Satisfaction with family health §	4	3.469	0.657	1	4	5052	Yes
(2) Satisfaction with public health workers §	3	3.291	0.791	1	4	4994	Yes
(3) Satisfaction with care ① ◇ §	4	3.658	0.670	1	4	2535	Yes
(4) Would return to clinic ◇ †	1	0.969	0.168	0	1	2527	Yes

†: dummy variable; §: likert scale (1-4); ¶: likert scale (1-5); ◇: measured across all health episode types (ante and post natal care, child birth, under-2 vaccinations, and illness and injury); ⓦ, ①, ⑥, and ⑫ stand for measures referring to the last week, one month, six months and twelve months respectively.

Table B.5: Clinic Quality

	Median	Mean	SD	Min	Max	N	BL data
(1) Clinic service provision index ①	0.034	0.068	1.143	−6.872	3.576	254	Yes
(2) Clinic aware of free health care †	1	0.803	0.398	0	1	254	Yes
(3) Employee satisfaction index	0.295	0.012	0.975	−3.652	1.443	254	Yes

†: dummy variable; §: likert scale (1-4); ¶: likert scale (1-5); ◇: measured across all health episode types (ante and post natal care, child birth, under-2 vaccinations, and illness and injury); ⓦ, ①, ⑥, and ⑫ stand for measures referring to the last week, one month, six months and twelve months respectively.

Table B.6: Health Service Delivery

	Median	Mean	SD	Min	Max	N	BL data
(1) Absenteeism index ① ◇	−0.239	0.057	1.070	−0.663	7.471	2874	Yes
(2) Paid for treatment ① ◇ †	0	0.404	0.491	0	1	2872	Yes
(3) Amount paid ① ◇	0	7816.945	35359.126	0	1360000	2843	Yes
(4) Any problem ① ◇ †	0	0.061	0.240	0	1	2869	Yes
(5) Staff not present ① ◇ †	0	0.020	0.140	0	1	2869	Yes
(6) Drugs not available ① ◇ †	0	0.027	0.162	0	1	2869	Yes
(7) Facility not clean ① ◇ †	0	0.002	0.042	0	1	2869	Yes
(8) Unpleasant staff behavior ① ◇ †	0	0.021	0.142	0	1	2869	Yes
(9) Medicine always in stock ① ◇ †	1	0.948	0.221	0	1	2478	No
(10) Individual satisfaction with care ① ◇ §	4	3.660	0.674	1	4	2863	Yes
(11) Individual would return to clinic ◇ †	1	0.971	0.165	0	1	2853	Yes
(12) Staff attitude ① ◇ §	4	3.752	0.555	1	4	2845	Yes

†: dummy variable; §: likert scale (1-4); ¶: likert scale (1-5); ◇: measured across all health episode types (ante and post natal care, child birth, under-2 vaccinations, and illness and injury); ⓦ, ①, ⑥, and ⑫ stand for measures referring to the last week, one month, six months and twelve months respectively.

Table B.7: Community Support

	Median	Mean	SD	Min	Max	N	BL data
(1) Contributions to clinic (community survey) ⑥	0.083	0.075	0.934	−1.170	4.193	508	Yes
(2) Contributions to clinic (facility survey) ⑥	0.317	−0.051	1.125	−5.373	1.503	508	Yes

In the community survey, we ask about meetings between the clinic and community as well as labor or financial contributions to the clinic. In the clinic survey, we ask about labor contributions as well as disputes between the community and the clinic. †: dummy variable; §: likert scale (1-4); ¶: likert scale (1-5); ◇: measured across all health episode types (ante and post natal care, child birth, under-2 vaccinations, and illness and injury); ⓦ, ①, ⑥, and ⑫ stand for measures referring to the last week, one month, six months and twelve months respectively.

Table B.8: Community development and political engagement (CDPE)

	Median	Mean	SD	Min	Max	N	BL data
(1) Projects with local council/chief ⑫ †	0	0.089	0.285	0	1	507	Yes
(2) Community provided labor ⑫ †	0	0.073	0.261	0	1	504	Yes
(3) Community involved in planning ⑫ †	0	0.050	0.217	0	1	504	Yes
(4) Problem addressed collectively? ⑫ †	1	0.573	0.495	0	1	508	Yes
(5) Proportion has voter card †	1	0.986	0.040	0.764	1	489	Yes
(6) Proportion voted in local election †	1	0.979	0.050	0.632	1	489	Yes
(7) Proportion voted in general election †	1	0.979	0.050	0.632	1	489	Yes

The Health Management Committee (HMC) meetings and the Facility Management Committee (FMC) meetings have been merged to a single meeting which we refer to as HMC. The Baseline variables for the government projects refers to projects in the past two years while at endline it was inquired for the past one year. †: dummy variable; §: likert scale (1-4); ¶: likert scale (1-5); ◇: measured across all health episode types (ante and post natal care, child birth, under-2 vaccinations, and illness and injury); ⓦ, ①, ⑥, and ⑫ stand for measures referring to the last week, one month, six months and twelve months respectively.

Table B.9: Water and Sanitation

	Median	Mean	SD	Min	Max	N	BL data
(1) Water and sanitation HH index	0.043	0.050	0.987	−5.667	6.236	5053	Yes
(2) Water and sanitation village index	−0.012	0.098	1.012	−1.212	3.791	5053	Yes
(3) Satisfaction with village sanitation index	0.290	0.050	0.973	−3.415	1.885	5051	Yes

The household index is comprised of water sources for drinking, for other uses, the existence and type of toilet facilities and the actions households take to make water safe to drink. The village index contains the existence and types of water sources for drinking and general use as well as the existence of public toilets and waste facilities. The satisfaction index consists of questions about sanitation and health services offered as well as village cleanliness. †: dummy variable; §: likert scale (1-4); ¶: likert scale (1-5); ◇: measured across all health episode types (ante and post natal care, child birth, under-2 vaccinations, and illness and injury); (w), (1), (6), and (12) stand for measures referring to the last week, one month, six months and twelve months respectively.

Table B.10: Economic Outcomes

	Median	Mean	SD	Min	Max	N	BL data
(1) Physical asset index	−0.245	0.014	1.052	−0.701	14.438	5052	Yes
(2) Agricultural asset index	−0.284	0.075	1.803	−0.560	58.153	5051	No
(3) Dwelling materials index	−0.163	0.038	1.018	−7.644	6.723	5052	Yes
(4) Total consumption expenditure (1)	−0.225	0.016	1.031	−1.311	14.540	5053	Yes

†: dummy variable; §: likert scale (1-4); ¶: likert scale (1-5); ◇: measured across all health episode types (ante and post natal care, child birth, under-2 vaccinations, and illness and injury); (w), (1), (6), and (12) stand for measures referring to the last week, one month, six months and twelve months respectively.

C.   Integrity of the experiment

Table C.1: Baseline Balance

	(1) Control Mean	(2) CM	(3) NFA	(4) Difference	(5) N
Village characteristics					
Motorable road	0.891	−0.009	0.005	−0.014	503
	(0.313)	(0.036)	(0.035)	(0.035)	
Mobile phone coverage	0.812	0.058	0.096	−0.038	504
	(0.392)	(0.044)	(0.041)**	(0.037)	
Distance to the closest clinic	1.362	−0.204	0.338	−0.542	504
	(2.217)	(0.329)	(0.481)	(0.463)	
Travel cost to closest clinic	94.225	−24.273	−24.389	0.116	503
	(869.677)	(72.811)	(74.502)	(65.453)	
Household characteristics and questions to household head					
Household size	3.369	−0.061	0.007	−0.068	4774
	(2.979)	(0.056)	(0.058)	(0.059)	
Number of illness or injury cases per household	0.054	−0.039	−0.026	−0.013	4774
	(0.237)	(0.011)***	(0.012)**	(0.013)	
Birth in household last year	0.157	−0.028	0.009	−0.037	2127
	(0.363)	(0.014)**	(0.015)	(0.014)**	
Child under 2 in household	0.230	−0.013	0.027	−0.040	2126
	(0.421)	(0.018)	(0.018)	(0.018)**	
Prominent village member in household	0.042	−0.007	−0.002	−0.005	2090
	(0.200)	(0.010)	(0.010)	(0.009)	
Believes doctor's advice	0.995	0.000	−0.007	0.007	1977
	(0.072)	(0.004)	(0.005)	(0.004)*	
Health care fees unaffordable	2.307	0.023	0.030	−0.007	2057
	(0.784)	(0.045)	(0.050)	(0.046)	
Trust in the community	1.856	−0.032	−0.010	−0.021	2127
	(0.663)	(0.051)	(0.049)	(0.052)	
Community cohesion	2.420	−0.018	0.009	−0.026	2122
	(0.610)	(0.034)	(0.036)	(0.035)	
Believe VHC members represent your interest	2.743	0.094	0.159	−0.066	984
	(1.061)	(0.107)	(0.122)	(0.104)	
The VHC can be trusted	2.453	−0.171	−0.089	−0.083	1148
	(0.967)	(0.103)*	(0.106)	(0.101)	
Individual characteristics					
Muslim	0.854	−0.037	−0.017	−0.021	9761
	(0.353)	(0.026)	(0.026)	(0.029)	
Mende (Ethnicity)	0.418	−0.019	−0.005	−0.014	9759
	(0.493)	(0.012)	(0.012)	(0.011)	
Temne (Ethnicity)	0.348	0.023	0.078	−0.055	9759
	(0.476)	(0.039)	(0.036)**	(0.035)	
Highest level of education	2.920	−0.013	0.256	−0.269	9734
	(4.203)	(0.124)	(0.134)*	(0.136)**	

Notes: This table illustrates the baseline balance across the three treatment arms (control, community monitoring and non-financial awards) among households surveyed in both waves. Column (1) shows the characteristics of the control group at baseline (mean and standard deviation). Columns (2) and (3) indicate the regression coefficients and standard errors of the CM and NFA treatment arms compared to the control group. Column (4) compares the two treatment arms. Household size and number of illness or injury cases per household contain a larger number of observations relative to other household level characteristics because these measures were also collected in the shorter user feedback survey discussed in Section 3.1.1. Significance: * is significant at the 10% level, ** is significant at the 5% level and *** is significant at the 1% level.

Table C.2: Manipulation Checks: Community Monitoring (CM)

	(1) Control Mean	(2) Pooled	(3) CM	(4) NFA	(5) N
(1) CM meetings by PLAN, Concern, or IRC took place	0.435	0.269	0.421	0.115	506
	(0.497)	(0.047)***	(0.050)***	(0.053)**	
(2) How many CM meetings took place?	0.897	1.055	1.631	0.483	498
	(1.218)	(0.142)***	(0.156)***	(0.157)***	
(3) Was village informed of meeting outcomes?	0.411	0.266	0.403	0.127	506
	(0.493)	(0.049)***	(0.054)***	(0.055)**	

Notes: Treatment effects are estimated using ANCOVA, controlling for the community-level average of the outcome family index at baseline and matching-triplet fixed effects. Robust standard errors, clustered by clinic, are shown in parentheses. Significance: * is significant at the 10% level, ** is significant at the 5% level and *** is significant at the 1% level.

Table C.3: Manipulation Checks: Non-Financial Awards (NFA)

	(1) Control Mean	(2) Pooled	(3) CM	(4) NFA	(5) N
(1) Clinic staff heard of NFA?	0.476	0.331	0.190	0.473	254
	(0.502)	(0.063)***	(0.073)***	(0.064)***	
(2) Clinic participated in NFA?	0.155	0.428	0.198	0.657	254
	(0.364)	(0.061)***	(0.067)***	(0.063)***	

Notes: Treatment effects are estimated using ANCOVA, controlling for the community-level average of the outcome family index at baseline and matching-triplet fixed effects. Robust standard errors are shown in parentheses. Significance: * is significant at the 10% level, ** is significant at the 5% level and *** is significant at the 1% level.

D. Endline results

D.1 Outcome Family tables (raw, not z-scored)

Table D.1: General Utilization

	(1) Control Mean	(2) Pooled	(3) CM	(4) NFA	(5) N
General utilization (SUR)		0.112 (0.031)*** [0.003]***	0.126 (0.034)*** [0.003]***	0.099 (0.036)*** [0.025]**	4496
(1) Number of health episodes in which sought western care	0.962 (0.393)	0.044 (0.012)*** [0.001]***	0.049 (0.013)*** [0.001]***	0.039 (0.014)*** [0.008]***	4496

Notes: Treatment effects are estimated using Missing-Indicator ANCOVA, controlling for the community-level average of the outcome family index at baseline and matching-triplet fixed effects. Robust standard errors, clustered by clinic, are shown in parentheses. Questions about ante and post natal care, vaccinations and illness/injury episodes are asked for the past one month. Questions about child birth are asked about the past year. Significance: * is significant at the 10% level, ** is significant at the 5% level and *** is significant at the 1% level.

Table D.2: Maternal Utilization

	(1) Control Mean	(2) Pooled	(3) CM	(4) NFA	(5) N
Maternal utilization index (SUR)		0.046 (0.045) [0.322]	0.130 (0.054)** [0.079]*	−0.031 (0.054) [0.463]	888
(1) ANC/PNC visits index	0.000 (1.000)	−0.038 (0.064) [0.386]	0.008 (0.082) [0.861]	−0.079 (0.068) [0.968]	887
(2) Birth in western medicine facility	0.834 (0.373)	0.048 (0.025)* [0.114]	0.094 (0.028)*** [0.002]***	0.006 (0.030) [0.968]	877

Notes: Treatment effects are estimated using Missing-Indicator ANCOVA, controlling for the community-level average of the outcome family index at baseline and matching-triplet fixed effects. Robust standard errors, clustered by clinic, are shown in parentheses. Multiple-inference corrected q-values that adjust for the false discovery rate within treatment arm are shown in square brackets. Significance: * is significant at the 10% level, ** is significant at the 5% level and *** is significant at the 1% level.

Table D.3: Health Outcomes

	(1) Control Mean	(2) Pooled	(3) CM	(4) NFA	(5) N
Health outcomes index (SUR)		0.021 (0.023) [0.322]	0.055 (0.025)** [0.094]*	−0.010 (0.027) [0.463]	5053
(1) U5 Child death per HH	0.039 (0.193)	−0.009 (0.005)* [0.544]	−0.015 (0.005)*** [0.039]**	−0.004 (0.006) [1.000]	5053
(2) Maternal death per HH	0.001 (0.035)	0.000 (0.001) [1.000]	0.001 (0.001) [0.462]	−0.001 (0.001) [1.000]	5053
(3) Illness/injury in HH	0.579 (0.494)	−0.003 (0.016) [1.000]	−0.009 (0.018) [0.462]	0.004 (0.019) [1.000]	5053
(4) Child weight for length	0.546 (1.682)	0.133 (0.081) [0.544]	0.156 (0.093)* [0.252]	0.109 (0.093) [1.000]	1991
(5) Vaccine completion index (Under 2)	3.085 (2.560)	0.032 (0.152) [1.000]	0.303 (0.184) [0.252]	−0.209 (0.163) [1.000]	1457
(6) Child birth complication index	0.000 (1.000)	−0.026 (0.077) [1.000]	−0.126 (0.086) [0.277]	0.061 (0.087) [1.000]	856
(7) Child illness index	0.002 (0.995)	0.024 (0.102) [1.000]	0.031 (0.115) [0.513]	0.018 (0.115) [1.000]	4993

Notes: Treatment effects are estimated using Missing-Indicator ANCOVA, controlling for the community-level average of the outcome family index at baseline and matching-triplet fixed effects. Robust standard errors, clustered by clinic, are shown in parentheses. Multiple-inference corrected q-values that adjust for the false discovery rate within treatment arm are shown in square brackets.Maternal death is defined as death relating to either pregnancy complications or childbirth. The Pentavalent Vaccination targets diphtheria, tetanus, whooping cough, hepatitis B, and haemophilus influenza type B. Significance: * is significant at the 10% level, ** is significant at the 5% level and *** is significant at the 1% level.

Table D.4: Satisfaction

	(1) Control Mean	(2) Pooled	(3) CM	(4) NFA	(5) N
Satisfaction index (SUR)		0.064 (0.030)** [0.058]*	0.059 (0.034)* [0.139]	0.069 (0.034)** [0.080]*	5052
(1) Satisfaction with family health	3.439 (0.670)	0.055 (0.027)** [0.096]*	0.051 (0.030)* [0.642]	0.059 (0.029)** [0.102]	5052
(2) Satisfaction with public health workers	3.258 (0.802)	0.069 (0.034)** [0.096]*	0.052 (0.041) [0.642]	0.083 (0.039)** [0.102]	4994
(3) Satisfaction with care	3.646 (0.696)	0.034 (0.035) [0.196]	0.039 (0.039) [0.642]	0.030 (0.039) [0.287]	2535
(4) Would return to clinic	0.967 (0.174)	0.007 (0.007) [0.196]	0.007 (0.008) [0.642]	0.007 (0.008) [0.287]	2527

Notes: Treatment effects are estimated using Missing-Indicator ANCOVA, controlling for the community-level average of the outcome family index at baseline and matching-triplet fixed effects. Robust standard errors, clustered by clinic, are shown in parentheses. Significance: * is significant at the 10% level, ** is significant at the 5% level and *** is significant at the 1% level.

Table D.5: Clinic Quality

	(1) Control Mean	(2) Pooled	(3) CM	(4) NFA	(5) N
Clinic quality index (SUR)		0.055 (0.064) [0.322]	−0.002 (0.074) [0.644]	0.112 (0.075) [0.157]	254
(1) Clinic service provision index	0.000 (1.000)	0.149 (0.153) [1.000]	0.095 (0.180) [1.000]	0.203 (0.175) [1.000]	254
(2) Clinic aware of free health care	0.798 (0.404)	0.005 (0.052) [1.000]	−0.023 (0.060) [1.000]	0.033 (0.058) [1.000]	254
(3) Employee satisfaction index	0.000 (1.000)	0.003 (0.127) [1.000]	−0.043 (0.141) [1.000]	0.049 (0.152) [1.000]	254

Notes: Treatment effects are estimated using ANCOVA, controlling for the community-level average of the outcome family index at baseline and matching-triplet fixed effects. Robust standard errors are shown in parentheses. Multiple-inference corrected q-values that adjust for the false discovery rate within treatment arm are shown in square brackets. The Clinic service provision index is composed of measures on facility maintenance (mainly cleanliness, orderly medicine storage and signposting) and whether required services like pre- and post-natal care, immunization, reproductive health and other forms of consultation are provided. The employee satisfaction index consists of employees' satisfaction with their job, with the communities' participation in the clinic and the extend to which they feel supported by the communities. Significance: * is significant at the 10% level, ** is significant at the 5% level and *** is significant at the 1% level.

Table D.6: Health Service Delivery

	(1) Control Mean	(2) Pooled	(3) CM	(4) NFA	(5) N
Health service delivery index (SUR)		0.021 (0.029) [0.355]	0.037 (0.035) [0.254]	0.015 (0.031) [0.463]	2877
(1) Absenteeism index	0.000 (1.000)	0.075 (0.057) [0.887]	−0.017 (0.070) [1.000]	0.109 (0.063)* [0.710]	2874
(2) Paid for treatment	0.416 (0.493)	−0.002 (0.023) [1.000]	−0.036 (0.033) [1.000]	0.011 (0.024) [0.786]	2872
(3) Amount paid	8520.158 (31895.827)	553.416 (1409.161) [0.948]	−1472.904 (1725.193) [1.000]	1321.751 (1496.821) [0.772]	2843
(4) Any problem	0.063 (0.242)	−0.000 (0.016) [1.000]	−0.002 (0.018) [1.000]	0.001 (0.017) [1.000]	2869
(5) Staff not present	0.019 (0.137)	0.005 (0.008) [0.887]	0.001 (0.010) [1.000]	0.006 (0.009) [0.772]	2869
(6) Drugs not available	0.031 (0.174)	−0.005 (0.009) [0.887]	−0.007 (0.010) [1.000]	−0.004 (0.010) [0.786]	2869
(7) Facility not clean	0.003 (0.058)	−0.002 (0.001) [0.887]	−0.003 (0.002) [1.000]	−0.002 (0.002) [0.710]	2869
(8) Unpleasant staff behavior	0.022 (0.148)	−0.010 (0.008) [0.887]	−0.005 (0.011) [1.000]	−0.013 (0.008) [0.710]	2869
(9) Medicine always in stock	0.952 (0.214)	0.006 (0.009) [0.887]	−0.001 (0.013) [1.000]	0.008 (0.010) [0.772]	2478
(10) Individual satisfaction with care	3.645 (0.703)	0.039 (0.039) [0.887]	0.069 (0.048) [1.000]	0.028 (0.041) [0.772]	2863
(11) Would return to clinic	0.969 (0.173)	0.013 (0.007)* [0.887]	0.018 (0.010)* [1.000]	0.011 (0.007) [0.710]	2853
(12) Staff attitude	3.735 (0.580)	0.043 (0.029) [0.887]	−0.011 (0.038) [1.000]	0.063 (0.031)** [0.710]	2845

Notes: Treatment effects are estimated using Missing-Indicator ANCOVA, controlling for the community-level average of the outcome family index at baseline and matching-triplet fixed effects. Robust standard errors, clustered by clinic, are shown in parentheses. Multiple-inference corrected q-values that adjust for the false discovery rate within treatment arm are shown in square brackets. The absenteeism index is composed of an indicator of whether patients had ever found no staff present when visiting the clinic and the waiting time at the last visit. Significance: * is significant at the 10% level, ** is significant at the 5% level and *** is significant at the 1% level.

Table D.7: Community Support

	(1) Control Mean	(2) Pooled	(3) CM	(4) NFA	(5) N
Contributions to clinic index (SUR)		0.025 (0.052) [0.465]	0.034 (0.061) [0.476]	0.016 (0.060) [0.463]	508
(1) Contributions to clinic index (community survey)	0.000 (1.000)	0.109 (0.078) [0.490]	0.134 (0.092) [0.404]	0.084 (0.092) [1.000]	508
(2) Contributions to clinic index (facility survey)	0.000 (0.997)	−0.059 (0.121) [0.490]	−0.067 (0.142) [0.471]	−0.052 (0.136) [1.000]	508

Notes: Treatment effects are estimated using ANCOVA, controlling for the community-level average of the outcome family index at baseline and matching-triplet fixed effects. Robust standard errors, clustered by clinic, are shown in parentheses. The two indices of contributions to the clinic are composed of variables that measure support and contributions once from the perspective of key informants in the villages and once by health personnel. In the community survey, we ask about meetings between the clinic and community as well as labor contributions to the clinic. In the clinic survey, we ask about labor or financial contributions as well as disputes between the community and the clinic. Significance: * is significant at the 10% level, ** is significant at the 5% level and *** is significant at the 1% level.

Table D.8: Community development and political engagement (CDPE)

	(1) Control Mean	(2) Pooled	(3) CM	(4) NFA	(5) N
CDPE index (SUR)		0.132 (0.049)*** [0.033]**	0.116 (0.060)* [0.114]	0.150 (0.056)*** [0.025]**	508
(1) Projects with local council/chief	0.054 (0.226)	0.047 (0.022)** [0.090]*	0.050 (0.026)* [0.465]	0.043 (0.026)* [0.129]	507
(2) Community provided labor	0.042 (0.201)	0.040 (0.019)** [0.090]*	0.038 (0.022)* [0.465]	0.042 (0.023)* [0.129]	504
(3) Community involved in planning	0.030 (0.171)	0.022 (0.016) [0.103]	0.026 (0.020) [0.465]	0.018 (0.020) [0.161]	504
(4) Problem addressed collectively?	0.583 (0.494)	−0.021 (0.040) [0.277]	−0.024 (0.045) [0.533]	−0.018 (0.046) [0.247]	508
(5) Proportion has voter card	0.983 (0.045)	0.005 (0.003) [0.103]	0.003 (0.004) [0.533]	0.007 (0.004)* [0.129]	489
(6) Proportion voted in local election	0.973 (0.054)	0.009 (0.004)** [0.090]*	0.005 (0.005) [0.465]	0.012 (0.005)** [0.099]*	489
(7) Proportion voted in general election	0.973 (0.055)	0.009 (0.005)** [0.090]*	0.008 (0.005) [0.465]	0.011 (0.005)** [0.109]	489

Notes: Treatment effects are estimated using ANCOVA, controlling for the community-level average of the outcome family index at baseline and matching-triplet fixed effects. Robust standard errors, clustered by clinic, are shown in parentheses. The Baseline variables for projects with the local council/chief refer to projects in the past two years while at endline it was inquired for the past one year. Significance: * is significant at the 10% level, ** is significant at the 5% level and *** is significant at the 1% level.

Table D.9: Water and Sanitation

	(1) Control Mean	(2) Pooled	(3) CM	(4) NFA	(5) N
Water and sanitation index (SUR)		0.110 (0.044)** [0.037]**	0.066 (0.051) [0.210]	0.154 (0.051)*** [0.025]**	5053
(1) Water and sanitation HH index	0.000 (1.000)	0.068 (0.058) [0.089]*	−0.022 (0.068) [0.362]	0.159 (0.063)** [0.040]**	5053
(2) Water and sanitation village index	−0.016 (1.003)	0.175 (0.087)** [0.073]*	0.132 (0.097) [0.298]	0.218 (0.104)** [0.040]**	5053
(3) Satisfaction with village sanitation index	0.000 (1.000)	0.087 (0.043)** [0.073]*	0.087 (0.049)* [0.298]	0.086 (0.050)* [0.059]*	5051

Notes: Treatment effects are estimated using Missing-Indicator ANCOVA, controlling for the community-level average of the outcome family index at baseline and matching-triplet fixed effects. Robust standard errors, clustered by clinic, are shown in parentheses. The household index is comprised of water sources for drinking, for other uses, the existence and type of toilet facilities and the actions households take to make water safe to drink. The village index contains the existence and types of water sources for drinking and general use as well as the existence of public toilets and waste facilities. The satisfaction index consists of questions about sanitation and health services offered as well as village cleanliness. Significance: * is significant at the 10% level, ** is significant at the 5% level and *** is significant at the 1% level.

Table D.10: Economic Outcomes

	(1) Control Mean	(2) Pooled	(3) CM	(4) NFA	(5) N
Economic outcomes index (SUR)		0.053 (0.032) [0.138]	0.036 (0.037) [0.262]	0.070 (0.040)* [0.119]	5053
(1) Physical asset index	0.000 (1.000)	0.020 (0.048) [1.000]	−0.023 (0.053) [1.000]	0.063 (0.059) [0.404]	5052
(2) Agricultural asset index	0.000 (1.000)	0.120 (0.044)*** [0.026]**	0.174 (0.057)*** [0.010]***	0.065 (0.050) [0.404]	5051
(3) Dwelling materials index	0.000 (1.000)	0.053 (0.053) [0.875]	0.020 (0.060) [1.000]	0.087 (0.058) [0.404]	5052
(4) Total consumption expenditure	0.000 (1.000)	0.019 (0.045) [1.000]	−0.027 (0.050) [1.000]	0.066 (0.055) [0.404]	5053

Notes: Treatment effects are estimated using Missing-Indicator ANCOVA, controlling for the community-level average of the outcome family index at baseline and matching-triplet fixed effects. Robust standard errors, clustered by clinic, are shown in parentheses. Significance: * is significant at the 10% level, ** is significant at the 5% level and *** is significant at the 1% level.

D.2 Outcome family tables (z-scored)

We omit the z-scored general utilization table here, since this outcome family has only one ingredient variable. The coefficients on that variable and its SUR index would replicate the coefficients reported in Table 1 and are therefore redundant.

Table D.11: Maternal utilization (z-scored)

	(1) Control Mean	(2) Pooled	(3) CM	(4) NFA	(5) N
Maternal utilization index (SUR)		0.046 (0.045) [0.322]	0.130 (0.054)** [0.079]*	−0.031 (0.054) [0.463]	888
(1) ANC/PNC visits index	0.000 (1.000)	−0.038 (0.064) [0.386]	0.008 (0.082) [0.861]	−0.079 (0.068) [0.968]	887
(2) Birth in western medicine facility	0.000 (1.000)	0.130 (0.066)* [0.114]	0.252 (0.074)*** [0.002]***	0.017 (0.082) [0.968]	877

Notes: Variables are control-group normalized at endline (z-scored). Treatment effects are estimated using Missing-Indicator ANCOVA, controlling for the community-level average of the outcome family index at baseline and matching-triplet fixed effects. Robust standard errors, clustered by clinic, are shown in parentheses. Multiple-inference corrected q-values that adjust for the false discovery rate within treatment arm are shown in square brackets. Significance: * is significant at the 10% level, ** is significant at the 5% level and *** is significant at the 1% level.

Table D.12: Health outcomes (z-scored)

	(1) Control Mean	(2) Pooled	(3) CM	(4) NFA	(5) N
Health outcomes index (SUR)		0.021 (0.023) [0.322]	0.055 (0.025)** [0.094]*	−0.010 (0.027) [0.463]	5053
(1) U5 Child death per HH	0.000 (1.000)	−0.047 (0.025)* [0.544]	−0.075 (0.027)*** [0.039]**	−0.019 (0.030) [1.000]	5053
(2) Maternal death per HH	0.000 (1.000)	0.000 (0.026) [1.000]	0.017 (0.031) [0.462]	−0.017 (0.028) [1.000]	5053
(3) Illness/injury in HH	0.000 (1.000)	−0.005 (0.033) [1.000]	−0.018 (0.037) [0.462]	0.007 (0.039) [1.000]	5053
(4) Child weight for length	0.000 (1.000)	0.079 (0.048) [0.544]	0.093 (0.055)* [0.252]	0.065 (0.055) [1.000]	1991
(5) Vaccine completion index (Under 2)	0.000 (1.000)	0.012 (0.059) [1.000]	0.118 (0.072) [0.252]	−0.082 (0.064) [1.000]	1457
(6) Child birth complication index	0.000 (1.000)	−0.026 (0.077) [1.000]	−0.126 (0.086) [0.277]	0.061 (0.087) [1.000]	856
(7) Child illness index	0.000 (1.000)	0.024 (0.102) [1.000]	0.031 (0.115) [0.513]	0.018 (0.116) [1.000]	4993

Notes: Variables are control-group normalized at endline (z-scored). Treatment effects are estimated using Missing-Indicator ANCOVA, controlling for the community-level average of the outcome family index at baseline and matching-triplet fixed effects. Robust standard errors, clustered by clinic, are shown in parentheses. Multiple-inference corrected q-values that adjust for the false discovery rate within treatment arm are shown in square brackets. Maternal death is defined as death relating to either pregnancy complications or childbirth. The Pentavalent Vaccination targets diphtheria, tetanus, whooping cough, hepatitis B, and haemophilus influenza type B. Significance: * is significant at the 10% level, ** is significant at the 5% level and *** is significant at the 1% level.

Table D.13: Satisfaction (z-scored)

	(1) Control Mean	(2) Pooled	(3) CM	(4) NFA	(5) N
Satisfaction index (SUR)		0.064 (0.030)** [0.058]*	0.059 (0.034)* [0.139]	0.069 (0.034)** [0.080]*	5052
(1) Satisfaction with family health	0.000 (1.000)	0.082 (0.040)** [0.096]*	0.076 (0.046)* [0.642]	0.088 (0.044)** [0.102]	5052
(2) Satisfaction with public health workers	0.000 (1.000)	0.085 (0.042)** [0.096]*	0.065 (0.051) [0.642]	0.103 (0.049)** [0.102]	4994
(3) Satisfaction with care	0.000 (1.000)	0.049 (0.050) [0.196]	0.056 (0.056) [0.642]	0.043 (0.056) [0.287]	2535
(4) Would return to clinic	0.000 (1.000)	0.039 (0.040) [0.196]	0.038 (0.048) [0.642]	0.041 (0.044) [0.287]	2527

Notes: Variables are control-group normalized at endline (z-scored). Treatment effects are estimated using Missing-Indicator ANCOVA, controlling for the community-level average of the outcome family index at baseline and matching-triplet fixed effects. Robust standard errors, clustered by clinic, are shown in parentheses. Significance: * is significant at the 10% level, ** is significant at the 5% level and *** is significant at the 1% level.

Table D.14: Clinic Quality (z-scored)

	(1) Control Mean	(2) Pooled	(3) CM	(4) NFA	(5) N
Clinic quality index (SUR)		0.055 (0.064) [0.322]	−0.002 (0.074) [0.644]	0.112 (0.075) [0.157]	254
(1) Clinic service provision index	0.000 (1.000)	0.149 (0.153) [1.000]	0.095 (0.180) [1.000]	0.203 (0.175) [1.000]	254
(2) Clinic aware of free health care	0.000 (1.000)	0.012 (0.129) [1.000]	−0.058 (0.148) [1.000]	0.083 (0.144) [1.000]	254
(3) Employee satisfaction index	0.000 (1.000)	0.003 (0.127) [1.000]	−0.043 (0.141) [1.000]	0.049 (0.152) [1.000]	254

Notes: Variables are control-group normalized at endline (z-scored). Treatment effects are estimated using ANCOVA, controlling for the community-level average of the outcome family index at baseline and matching-triplet fixed effects. Robust standard errors are shown in parentheses. Multiple-inference corrected q-values that adjust for the false discovery rate within treatment arm are shown in square brackets. The Clinic service provision index is composed of measures on facility maintenance (mainly cleanliness, orderly medicine storage and signposting) and whether required services like pre- and post-natal care, immunization, reproductive health and other forms of consultation are provided. The employee satisfaction index consists of employees' satisfaction with their job, with the communities' participation in the clinic and the extend to which they feel supported by the communities. Significance: * is significant at the 10% level, ** is significant at the 5% level and *** is significant at the 1% level.

Table D.15: Health Service Delivery (z-scored)

	(1) Control Mean	(2) Pooled	(3) CM	(4) NFA	(5) N
Health service delivery index (SUR)		0.021	0.037	0.015	2877
		(0.029)	(0.035)	(0.031)	
		[0.355]	[0.254]	[0.463]	
(1) Absenteeism index	0.000	0.075	−0.017	0.109	2874
	(1.000)	(0.057)	(0.070)	(0.063)*	
		[0.887]	[1.000]	[0.710]	
(2) Paid for treatment	0.000	−0.004	−0.074	0.022	2872
	(1.000)	(0.047)	(0.066)	(0.048)	
		[1.000]	[1.000]	[0.786]	
(3) Amount paid	0.000	0.017	−0.046	0.041	2843
	(1.000)	(0.044)	(0.054)	(0.047)	
		[0.948]	[1.000]	[0.772]	
(4) Any problem	0.000	−0.000	−0.008	0.003	2869
	(1.000)	(0.066)	(0.074)	(0.069)	
		[1.000]	[1.000]	[1.000]	
(5) Staff not present	0.000	0.035	0.004	0.047	2869
	(1.000)	(0.062)	(0.070)	(0.066)	
		[0.887]	[1.000]	[0.772]	
(6) Drugs not available	0.000	−0.030	−0.042	−0.025	2869
	(1.000)	(0.052)	(0.056)	(0.056)	
		[0.887]	[1.000]	[0.786]	
(7) Facility not clean	0.000	−0.042	−0.053	−0.037	2869
	(1.000)	(0.026)	(0.034)	(0.028)	
		[0.887]	[1.000]	[0.710]	
(8) Unpleasant staff behavior	0.000	−0.071	−0.033	−0.085	2869
	(1.000)	(0.054)	(0.072)	(0.053)	
		[0.887]	[1.000]	[0.710]	
(9) Medicine always in stock	0.000	0.026	−0.006	0.038	2478
	(1.000)	(0.043)	(0.059)	(0.045)	
		[0.887]	[1.000]	[0.772]	
(10) Individual satisfaction with care	0.000	0.056	0.098	0.041	2863
	(1.000)	(0.056)	(0.068)	(0.059)	
		[0.887]	[1.000]	[0.772]	
(11) Would return to clinic	0.000	0.074	0.105	0.063	2853
	(1.000)	(0.042)*	(0.057)*	(0.042)	
		[0.887]	[1.000]	[0.710]	
(12) Staff attitude	0.000	0.075	−0.019	0.109	2845
	(1.000)	(0.050)	(0.066)	(0.053)**	
		[0.887]	[1.000]	[0.710]	

Notes: Variables are control-group normalized at endline (z-scored). Treatment effects are estimated using Missing-Indicator ANCOVA, controlling for the community-level average of the outcome family index at baseline and matching-triplet fixed effects. Robust standard errors, clustered by clinic, are shown in parentheses. Multiple-inference corrected q-values that adjust for the false discovery rate within treatment arm are shown in square brackets.The absenteeism index is composed of an indicator of whether patients had ever found no staff present when visiting the clinic and the waiting time at the last visit. Significance: * is significant at the 10% level, ** is significant at the 5% level and *** is significant at the 1% level.

Table D.16: Contributions to Clinic (z-scored)

	(1) Control Mean	(2) Pooled	(3) CM	(4) NFA	(5) N
Contributions to clinic index (SUR)		0.025 (0.052) [0.465]	0.034 (0.061) [0.476]	0.016 (0.060) [0.463]	508
(1) Contributions to clinic index (community survey)	0.000 (1.000)	0.109 (0.078) [0.490]	0.134 (0.092) [0.404]	0.084 (0.092) [1.000]	508
(2) Contributions to clinic index (facility survey)	0.000 (1.000)	−0.060 (0.121) [0.490]	−0.067 (0.143) [0.471]	−0.052 (0.136) [1.000]	508

Notes: Variables are control-group normalized at endline (z-scored). Treatment effects are estimated using ANCOVA, controlling for the community-level average of the outcome family index at baseline and matching-triplet fixed effects. Robust standard errors, clustered by clinic, are shown in parentheses. The two indices of contributions to the clinic are composed of variables that measure support and contributions once from the perspective of key informants in the villages and once by health personnel. In the community survey, we ask about meetings between the clinic and community as well as labor contributions to the clinic. In the clinic survey, we ask about labor or financial contributions as well as disputes between the community and the clinic. Significance: * is significant at the 10% level, ** is significant at the 5% level and *** is significant at the 1% level.

Table D.17: Community Development and Political Engagement (CDPE, z-scored)

	(1) Control Mean	(2) Pooled	(3) CM	(4) NFA	(5) N
CDPE index (SUR)		0.132 (0.049)*** [0.033]**	0.116 (0.060)* [0.114]	0.150 (0.056)*** [0.025]**	508
(1) Projects with local council/chief	0.000 (1.000)	0.208 (0.095)** [0.090]*	0.223 (0.114)* [0.465]	0.192 (0.116)* [0.129]	507
(2) Community provided labor	0.000 (1.000)	0.200 (0.095)** [0.090]*	0.190 (0.112)* [0.465]	0.209 (0.116)* [0.129]	504
(3) Community involved in planning	0.000 (1.000)	0.127 (0.095) [0.103]	0.150 (0.117) [0.465]	0.103 (0.115) [0.161]	504
(4) Problem addressed collectively?	0.000 (1.000)	−0.043 (0.080) [0.277]	−0.049 (0.092) [0.533]	−0.037 (0.094) [0.247]	508
(5) Proportion has voter card	0.000 (1.000)	0.102 (0.076) [0.103]	0.061 (0.087) [0.533]	0.145 (0.087)* [0.129]	489
(6) Proportion voted in local election	0.000 (1.000)	0.163 (0.082)** [0.090]*	0.099 (0.094) [0.465]	0.231 (0.092)** [0.099]*	489
(7) Proportion voted in general election	0.000 (1.000)	0.172 (0.085)** [0.090]*	0.140 (0.097) [0.465]	0.206 (0.096)** [0.109]	489

Notes: Variables are control-group normalized at endline (z-scored). Treatment effects are estimated using ANCOVA, controlling for the community-level average of the outcome family index at baseline and matching-triplet fixed effects. Robust standard errors, clustered by clinic, are shown in parentheses. The Baseline variables for projects with the local council/chief refer to projects in the past two years while at endline it was inquired for the past one year. Significance: * is significant at the 10% level, ** is significant at the 5% level and *** is significant at the 1% level.

Table D.18: Water and Sanitation (z-scored)

	(1) Control Mean	(2) Pooled	(3) CM	(4) NFA	(5) N
Water and sanitation index (SUR)		0.110 (0.044)** [0.037]**	0.066 (0.051) [0.210]	0.154 (0.051)*** [0.025]**	5053
(1) Water and sanitation HH index	0.000 (1.000)	0.068 (0.058) [0.089]*	−0.022 (0.068) [0.362]	0.159 (0.063)** [0.040]**	5053
(2) Water and sanitation village index	0.000 (1.000)	0.174 (0.086)** [0.073]*	0.131 (0.097) [0.298]	0.218 (0.104)** [0.040]**	5053
(3) Satisfaction with village sanitation index	0.000 (1.000)	0.087 (0.043)** [0.073]*	0.087 (0.049)* [0.298]	0.086 (0.050)* [0.059]*	5051

Notes: Variables are control-group normalized at endline (z-scored). Treatment effects are estimated using Missing-Indicator ANCOVA, controlling for the community-level average of the outcome family index at baseline and matching-triplet fixed effects. Robust standard errors, clustered by clinic, are shown in parentheses. The household index is comprised of water sources for drinking, for other uses, the existence and type of toilet facilities and the actions households take to make water safe to drink. The village index contains the existence and types of water sources for drinking and general use as well as the existence of public toilets and waste facilities. The satisfaction index consists of questions about sanitation and health services offered as well as village cleanliness. Significance: * is significant at the 10% level, ** is significant at the 5% level and *** is significant at the 1% level.

Table D.19: Economic Outcomes (z-scored)

	(1) Control Mean	(2) Pooled	(3) CM	(4) NFA	(5) N
Economic outcomes index (SUR)		0.053 (0.032) [0.138]	0.036 (0.037) [0.262]	0.070 (0.040)* [0.119]	5053
(1) Physical asset index	0.000 (1.000)	0.020 (0.048) [1.000]	−0.023 (0.053) [1.000]	0.063 (0.059) [0.404]	5052
(2) Agricultural asset index	0.000 (1.000)	0.120 (0.044)*** [0.026]**	0.174 (0.057)*** [0.010]***	0.065 (0.050) [0.404]	5051
(3) Dwelling materials index	0.000 (1.000)	0.053 (0.053) [0.875]	0.020 (0.060) [1.000]	0.087 (0.058) [0.404]	5052
(4) Total consumption expenditure	0.000 (1.000)	0.019 (0.045) [1.000]	−0.027 (0.050) [1.000]	0.066 (0.055) [0.404]	5053

Notes: Variables are control-group normalized at endline (z-scored). Treatment effects are estimated using Missing-Indicator ANCOVA, controlling for the community-level average of the outcome family index at baseline and matching-triplet fixed effects. Robust standard errors, clustered by clinic, are shown in parentheses. Significance: * is significant at the 10% level, ** is significant at the 5% level and *** is significant at the 1% level.

D.3 Additional outcome tables

Table D.20: Perceived quality of care

	Control Mean	Pooled	CM	NFA	N
(1) **Perceived quality of care index**	0.000	0.351	0.363	0.340	254
	(1.000)	(0.121)***	(0.139)***	(0.143)**	
(2) General utilization	0.962	0.044	0.049	0.039	4496
	(0.393)	(0.012)***	(0.013)***	(0.014)***	
(3) Satisfaction with public health workers	3.258	0.056	0.036	0.076	4994
	(0.802)	(0.034)*	(0.040)	(0.039)*	
(4) Effectiveness of western medicine relative to traditional or religious healing	−0.361	0.036	0.045	0.028	5053
	(0.200)	(0.020)*	(0.023)*	(0.023)	

Notes: Treatment effects are estimated using Missing-Indicator ANCOVA, controlling for the community-level average of the outcome family index at baseline and matching-triplet fixed effects. Robust standard errors, clustered by clinic, are shown in parentheses. Significance: * is significant at the 10% level, ** is significant at the 5% level and *** is significant at the 1% level.

Table D.21: PHU utilization

	Control Mean	Pooled	CM	NFA	N
(1) Number of health episodes in which sought western care	0.962	0.044	0.049	0.039	4496
	(0.393)	(0.012)***	(0.013)***	(0.014)***	
(2) Number of health episodes in which sought western care at PHU	0.883	0.059	0.065	0.053	4496
	(0.463)	(0.017)***	(0.019)***	(0.020)**	

Notes: Treatment effects are estimated using Missing-Indicator ANCOVA, controlling for the community-level average of the outcome family index at baseline and matching-triplet fixed effects. Robust standard errors, clustered by clinic, are shown in parentheses. Significance: * is significant at the 10% level, ** is significant at the 5% level and *** is significant at the 1% level.

Table D.22: Comparing Effects on Two Absenteeism Measures

	(1) Control Mean	(2) Pooled	(3) CM	(4) NFA	(5) N
(1) Ever no staff present among all clinic visits	0.055	0.034	0.021	0.048	2870
	(0.228)	(0.012)***	(0.014)	(0.015)***	
(2) No staff present on last clinic visit	0.007	0.005	0.007	0.004	1885
	(0.083)	(0.006)	(0.006)	(0.006)	

Notes: Treatment effects are estimated using Missing-Indicator ANCOVA, controlling for the community-level average of the outcome family index at baseline and matching-triplet fixed effects. Robust standard errors, clustered by clinic, are shown in parentheses. Significance: * is significant at the 10% level, ** is significant at the 5% level and *** is significant at the 1% level.

Table D.23: Main outcome families when controlling for baseline imbalances

	(1) Control Mean	(2) Pooled	(3) CM	(4) NFA	(5) N
General utilization	0.000	0.097	0.106	0.087	4451
	(1.000)	(0.033)***	(0.035)***	(0.039)**	
		[0.036]**	[0.014]**	[0.080]*	
Maternal utilization index	0.000	0.060	0.178	−0.059	878
	(1.000)	(0.070)	(0.081)**	(0.082)	
		[0.277]	[0.083]*	[0.473]	
Health outcomes index	0.000	0.082	0.170	−0.010	5003
	(1.000)	(0.050)	(0.056)***	(0.058)	
		[0.188]	[0.014]**	[0.909]	
Satisfaction index	0.000	0.110	0.094	0.124	5002
	(1.000)	(0.044)**	(0.051)*	(0.049)**	
		[0.060]*	[0.134]	[0.072]*	
Health service delivery index	0.000	0.055	0.082	0.044	2845
	(1.000)	(0.061)	(0.074)	(0.063)	
		[0.277]	[0.330]	[0.473]	
Clinic quality index	0.000	0.151	0.057	0.260	254
	(1.000)	(0.157)	(0.183)	(0.184)	
		[0.277]	[0.606]	[0.290]	
CDPE index	0.000	0.161	0.153	0.168	501
	(1.000)	(0.086)*	(0.104)	(0.101)*	
		[0.145]	[0.203]	[0.208]	
Contributions to clinic index	0.000	0.024	0.044	0.002	501
	(1.000)	(0.097)	(0.113)	(0.111)	
		[0.678]	[0.606]	[0.969]	
Water and sanitation index	0.000	0.122	0.069	0.177	5003
	(1.000)	(0.062)**	(0.071)	(0.071)**	
		[0.145]	[0.330]	[0.072]*	
Economic outcomes index	0.000	0.057	0.063	0.051	5003
	(1.000)	(0.051)	(0.060)	(0.061)	
		[0.277]	[0.330]	[0.473]	

Notes: Treatment effects are estimated using Missing-Indicator ANCOVA, controlling for the community-level average of the outcome family index at baseline and matching-triplet fixed effects. Robust standard errors, clustered by clinic, are shown in parentheses. Multiple-inference corrected q-values that adjust for the false discovery rate within treatment arm are shown in square brackets. We also control for baseline variables displaying imbalance in Table C.1 – namely, phone coverage, household size, the number of births in the household in the last year, the share of the village population of Temne ethnicity, the highest level of educational attainment, whether they believe what the doctors tell them, and the number of illness or injury cases in the household. Where imbalanced baseline characteristics are measured at a different level of observation, we average or assign to each member as needed. Significance: * is significant at the 10% level, ** is significant at the 5% level and *** is significant at the 1% level.

E. Ebola

E.1 Specialized Ebola Facilities and Training

The UN Mission for Emergency Ebola Response (UNMEER) compiled information on three types of treatment facilities:

1. Ebola Treatment Unit (ETU): 17 facilities with an average of 32 beds;

2. Ebola Holding Center (EHC): 49 facilities with an average of 18 beds; and

3. Community Care Center (CCC): 41 facilities with an average of 10 beds

Figure E.1: Location of Ebola Treatment Facilities

(a) UNMEER (b) National Ebola Response Center (NERC)
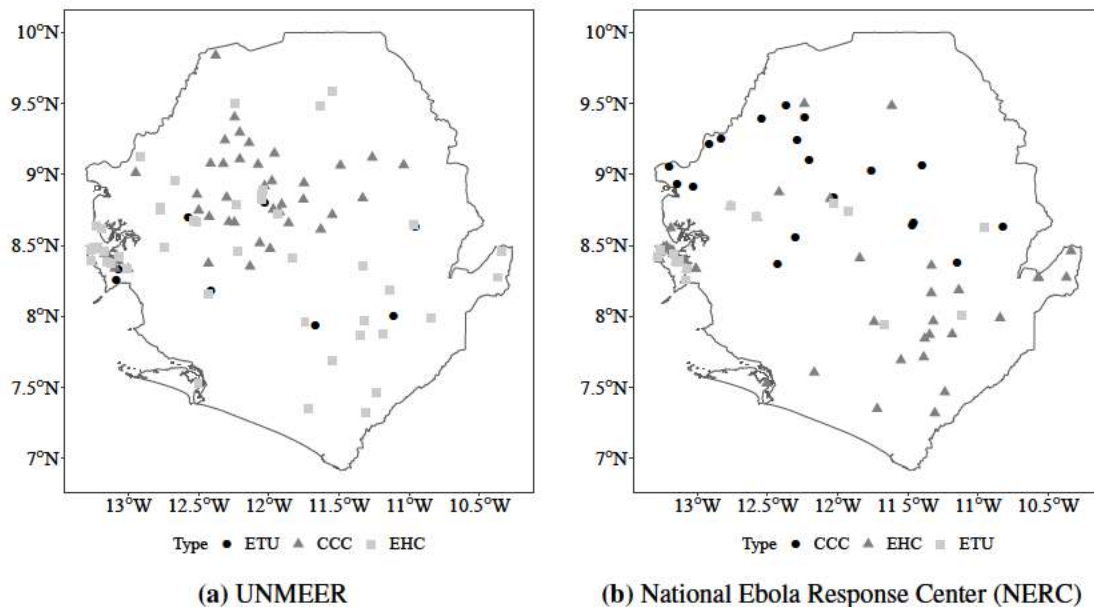
Maps of three types of treatment facilities: Ebola Treatment Units (ETUs), Ebola Holding Centers (EHCs), and Community Care Centers (CCCs). The plots differ in the source of the information: data on the left come from the National Ebola Response Center (NERC); the right, from UNMEER. These sources largely overlap, though the NERC data contains fewer CCCs and more missing geo-coordinates than the UNMEER data. Both datasets were accessed through the Humanitarian Data Exchange.

Table E.1: Average Reported Cases in Sections with Specialized Ebola Facilities

Facility Type	ETU	EHC	CCC	Any
Total Cases				
No Facility	25.2	26.6	27.9	22.6
Facility Present	484.0	126.5	99.6	131.3
Confirmed Cases				
No Facility	3.2	3.3	3.6	2.8
Facility Present	62.6	18.5	11.3	17.8
Negative Cases				
No Facility	18.1	19.1	20.0	16.2
Facility Present	340.0	92.4	74.0	94.9

Notes: Data on facility locations taken from UNMEER.

Table E.2: Health Care Worker (HCW) Training Schedule

Week Ending	HCWs Trained	% Total HCWs Trained (4,264)
11/28/2014	2,440	57%
12/05/2014	3,450	81%
12/12/2014	3,980	93%
12/19/2014	4,200	98%
12/26/2014	4,200	98%

Notes: Approximate counts extracted from report, "Infection Prevention and Control (IPC) and Screening of Suspected Ebola Cases," p. 4.

E.2 Effect on Reported Cases by Month

Table E.3: Effect on Total Cases

	Control Mean	Pooled	CM	NFA	Difference	N
09-07 to 09-28	0.093	0.064	0.096	0.04	0.057	640
	(0.432)	(0.051)	(0.085)	(0.079)	(0.128)	
10-05 to 10-26	0.065	0.325	0.393	0.272	0.122	640
	(0.298)	(0.132)**	(0.201)*	(0.141)*	(0.217)	
11-02 to 11-30	0.248	0.16	0.16	0.161	-0.001	800
	(0.733)	(0.114)	(0.155)	(0.128)	(0.165)	
12-07 to 12-28	0.167	0.216	0.109	0.298	-0.189	640
	(0.472)	(0.087)**	(0.109)	(0.117)**	(0.147)	
01-04 to 02-01	0.485	0.055	-0.009	0.105	-0.114	800
	(0.932)	(0.109)	(0.144)	(0.127)	(0.159)	
02-08 to 03-01	0.319	0.287	0.337	0.249	0.088	640
	(0.804)	(0.158)*	(0.237)	(0.169)	(0.252)	
03-08 to 03-29	0.495	0.091	0.231	-0.017	0.248	640
	(0.894)	(0.133)	(0.184)	(0.151)	(0.206)	
04-05 to 04-26	0.329	0.214	0.376	0.089	0.287	640
	(0.783)	(0.113)*	(0.151)**	(0.125)	(0.16)*	
05-03 to 05-10	0.528	0.03	0.062	0.005	0.057	320
	(1.045)	(0.172)	(0.212)	(0.21)	(0.245)	

Notes: Standard errors clustered on section. Treatment effects estimated using OLS including matching-triplet and week fixed effects. N varies because case counts are recorded weekly, and there can be 4 or 5 weeks within each period. Significance: *p < 0.10, ** p < 0.05, and *** p < 0.01.

E.3 Time-series of Confirmed Cases

Figure E.2: Confirmed Cases by Treatment

(a) Weekly Counts

(b) Cumulative Counts

Figure E.2(a) plots the time series of confirmed cases by week; bars represent the raw counts. We use the date that the case was first saved in the VHF, which is available for 96 percent of cases in our sample. Figure E.2(b) graphs the cumulative count of confirmed cases by treatment group.

E.4 Effect on Patient Deaths

Table E.4: Effect on Patient Deaths

	Dependent variable:	
	Patient Deaths	
	(1)	(2)
Total Cases in Last 2 Weeks	0.245	0.247
	$(0.021)^{***}$	$(0.021)^{***}$
Pooled	0.063	
	$(0.032)^{**}$	
Total Cases in Last 2 Weeks $\times$ Pooled	-0.098	
	$(0.043)^{**}$	
CM		0.116
		$(0.037)^{***}$
Total Cases in Last 2 Weeks $\times$ CM		-0.149
		$(0.046)^{***}$
NFA		-0.007
		(0.025)
Total Cases in Last 2 Weeks $\times$ NFA		-0.019
		(0.032)
Control Mean	0.149	0.149
	(0.49)	(0.49)
Observations	5,280	5,280

Notes: Standard errors clustered on section. Treatment effects estimated using OLS including matching-triplet and week fixed effects. Significance: * $p < 0.10$, ** $p < 0.05$, and *** $p < 0.01$.

Table E.5: Effect on Delays between Symptom Onset and Reporting

	Control Mean	Pooled	CM	NFA	Difference	N
Delay: Symptom Onset and Reporting	4.729	0.218	0.583	-0.066	0.649	160
	(3.229)	(0.51)	(0.62)	(0.579)	(0.626)	

Notes: Treatment effects estimated using OLS including matching-triplet fixed effects. Significance: *$p < 0.10$, ** $p < 0.05$, and *** $p < 0.01$. Delays greater than 60 days were removed to limit the influence of outliers.

E.5   Effect on Probable and Suspected Cases

During the study period from September 2014 until April 2015, there are 19 probable and 134 suspected cases. The VHF uses the following criteria to classify probable and suspected cases:

- **Probable** (unconfirmed) cases are suspected cases that meet one of two additional criteria: (1) they were screened by a clinician; or (2) deceased individuals with an epidemiological link with a confirmed case. In our sample and study period, there are 19 probable cases.

- **Suspected** cases include (1) the onset of high fever and contact with a suspected, probable, or confirmed individuals or a dead or sick animal; (2) the onset of high fever and at least three of the following symptoms: headaches, vomiting, anorexia/loss of appetite, diarrhea, lethargy, stomach pain, aching muscles or joints, difficulty swallowing, breathing difficulties, or hiccup; any person with inexplicable bleeding; or any sudden, inexplicable death. In our sample and study period, there are 134 suspected cases.

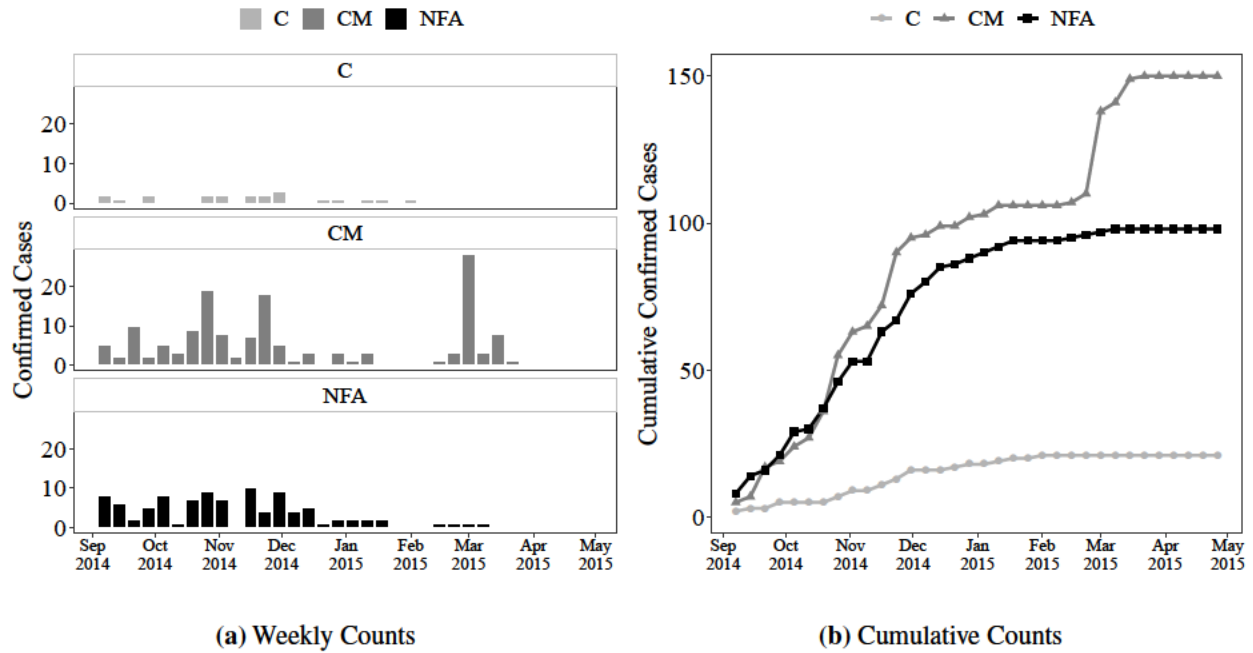**Figure E.3:** Weekly Counts of Probable and Suspected Cases by Treatment
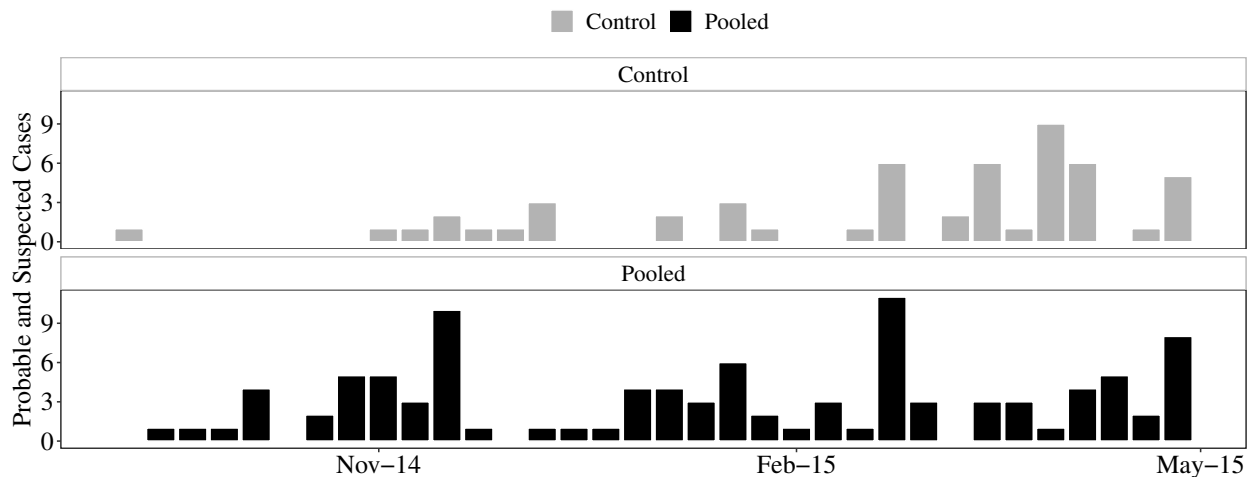
The time series of probable and suspected cases by week; bars represent the raw counts. We use the date that the case was first saved in the VHF.

Table E.6: Effect on Probable and Suspected Cases

	Control Mean	Pooled	CM	NFA	Difference	N
Ebola Cases						
Probable and Suspected	0.029	0.003	0.015	-0.007	0.022	5,440
	(0.208)	(0.008)	(0.011)	(0.009)	(0.011)*	
Log(Ebola Cases + 1)						
Probable and Suspected	0.018	0.002	0.009	-0.004	0.013	5,440
	(0.123)	(0.005)	(0.007)	(0.005)	(0.007)*	
Linear Probability Model						
Probable and Suspected	0.022	0.002	0.011	-0.005	0.016	5,440
	(0.148)	(0.006)	(0.008)	(0.006)	(0.008)*	
IHS(Ebola Cases)						
Probable and Suspected	0.023	0.002	0.012	-0.005	0.017	5,440
	(0.159)	(0.006)	(0.008)	(0.007)	(0.009)*	

Notes: Standard errors clustered on section shown in parentheses. Treatment effects estimated using OLS including matching-triplet and week fixed effects. Significance: *$p < 0.10$, ** $p < 0.05$, and *** $p < 0.01$.

Table E.7: Effect on Reported Cases (Removing Probable and Suspected)

	Control Mean	Pooled	CM	NFA	Difference	N
Ebola Cases						
Total	0.252	0.17	0.188	0.155	0.033	5,440
	(0.672)	(0.08)**	(0.11)*	(0.094)	(0.127)	
IHS(Ebola Cases)						
Total	1.004	0.057	0.061	0.053	0.007	5,440
	(0.292)	(0.028)**	(0.037)*	(0.033)	(0.042)	

Notes: The sample includes 160 sections over 34 weeks. Standard errors clustered on section shown in parentheses. Treatment effects estimated using OLS including matching-triplet and week fixed effects. Difference column reports the difference between the CM and NFA coefficients; the standard error is computed using the delta method. Significance: * $p < 0.10$, ** $p < 0.05$, and *** $p < 0.01$.

E.6 Dropping Triplets

Figure E.4: Estimates Dropping Each Triplet from Sample

(a) Coefficient $(\widehat{\beta})$
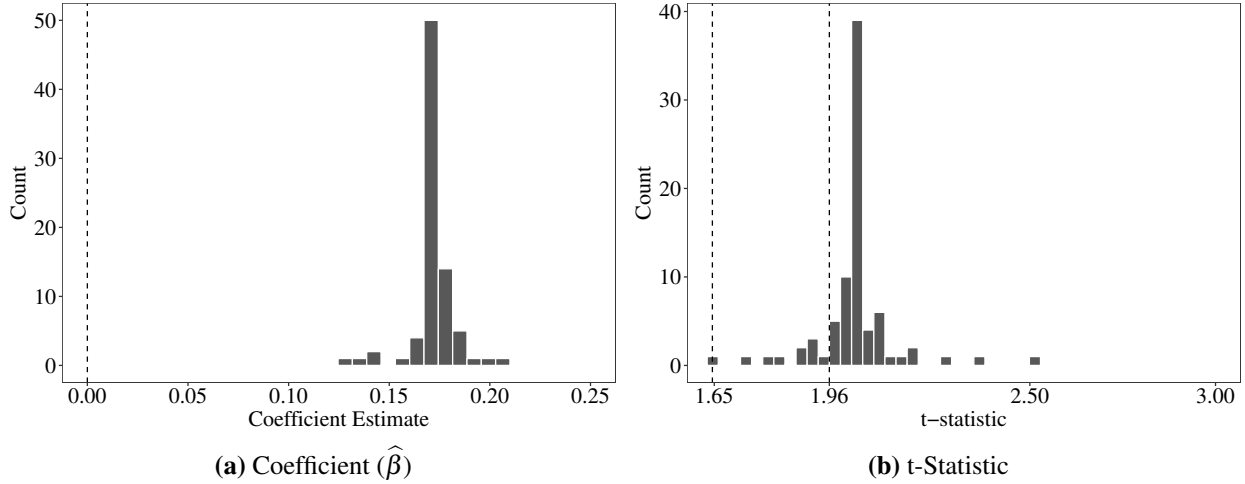
(b) t-Statistic

We re-estimate Equation 3 dropping one triplet (i.e., block) from the sample with each iteration. Figure E.4(a): distribution of coefficient estimates. Figure E.4(b): distribution of t-statistics.

Figure E.5: Estimates Dropping Each Pair of Triplets from Sample

(a) Coefficient $(\widehat{\beta})$
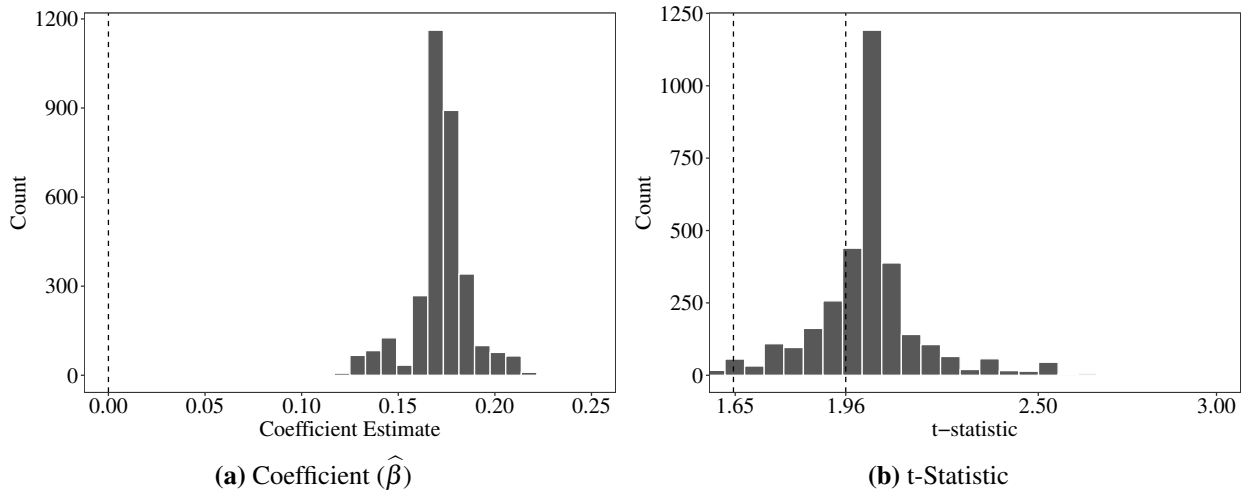
(b) t-Statistic

We re-estimate Equation 3 dropping pairs of triplets (i.e., block) from the sample with each iteration. Figure E.4(a): distribution of coefficient estimates. Figure E.4(b): distribution of t-statistics.

E.7 Dropping Weeks

Figure E.6: Estimates Dropping Each Week from Sample

(a) Coefficient ($\widehat{\beta}$)
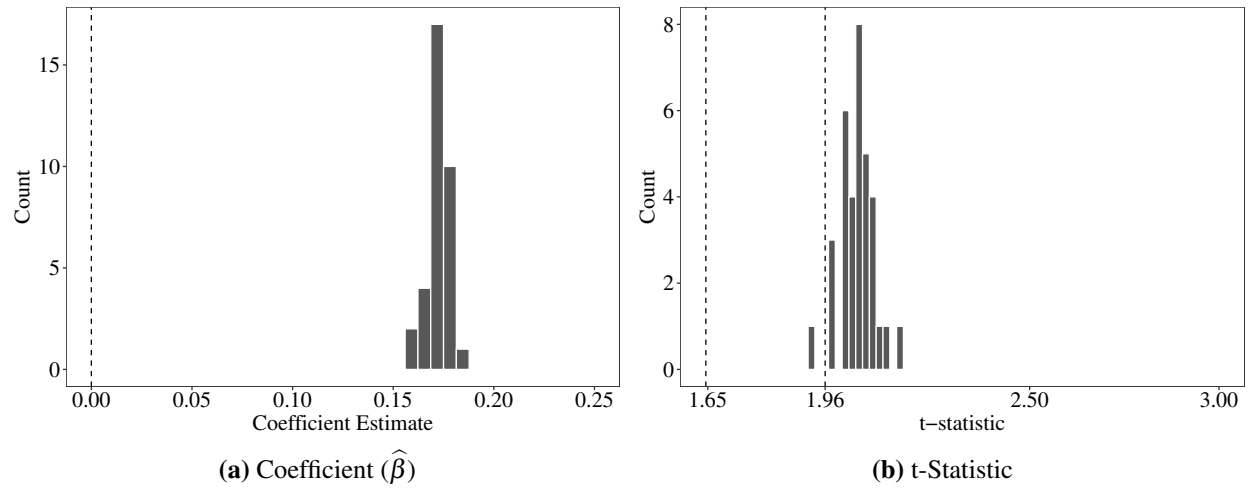
(b) t-Statistic

We re-estimate Equation 3 dropping one week from the sample with each iteration. Figure E.6(a): distribution of coefficient estimates. Figure E.6(b): distribution of t-statistics.

E.8 Alternative Functional Forms for Reported Cases

Table E.8: Effect on Reported Cases (Alternative Specifications)

	Control Mean	Pooled	CM	NFA	Difference	N
Linear Probability Model						
Total	0.182	0.047	0.057	0.039	0.018	5,440
	(0.386)	(0.026)*	(0.034)*	(0.031)	(0.038)	
Confirmed	0.009	0.021	0.022	0.02	0.002	5,440
	(0.096)	(0.007)***	(0.01)**	(0.008)***	(0.011)	
Negative	0.164	0.038	0.038	0.037	0.002	5,440
	(0.371)	(0.024)	(0.03)	(0.029)	(0.034)	
Log(Ebola Cases + 1)						
Total	0.16	0.065	0.074	0.057	0.017	5,440
	(0.363)	(0.033)*	(0.044)*	(0.039)	(0.05)	
Confirmed	0.007	0.023	0.027	0.019	0.008	5,440
	(0.077)	(0.008)***	(0.012)**	(0.009)**	(0.013)	
Negative	0.139	0.045	0.04	0.049	-0.008	5,440
	(0.335)	(0.027)*	(0.034)	(0.033)	(0.04)	
Poisson						
Total	0.281	0.469	0.552	0.407	0.144	5,440
	(0.727)	(0.213)**	(0.286)*	(0.25)	(0.322)	
Confirmed	0.011	1.669	2.008	1.369	0.639	5,440
	(0.129)	(0.513)***	(0.58)***	(0.569)**	(0.5)	
Negative	0.238	0.342	0.276	0.389	-0.113	5,440
	(0.648)	(0.201)*	(0.268)	(0.234)*	(0.3)	

Notes: Standard errors clustered on section. Treatment effects estimated using OLS including matching-triplet and week fixed effects. Significance: * p < 0.10, ** p < 0.05, and *** p < 0.01

E.9 Extending Panel to August 2014

Table E.9: Effect on Reported Cases in Extended Panel (August 2014–April 2015)

	Control Mean	Pooled	CM	NFA	Difference	N
Ebola Cases						
Total	0.257	0.163	0.189	0.142	0.048	6,079
	(0.706)	(0.077)**	(0.106)*	(0.091)	(0.122)	
Confirmed	0.014	0.058	0.079	0.041	0.038	6,079
	(0.146)	(0.022)**	(0.034)**	(0.025)*	(0.038)	
Negative	0.216	0.092	0.074	0.105	-0.031	6,079
	(0.621)	(0.055)*	(0.07)	(0.068)	(0.084)	
IHS(Ebola Cases)						
Total	0.189	0.078	0.088	0.07	0.018	6,079
	(0.454)	(0.039)**	(0.052)*	(0.046)	(0.059)	
Confirmed	0.011	0.028	0.032	0.025	0.007	6,079
	(0.109)	(0.01)***	(0.014)**	(0.011)**	(0.016)	
Negative	0.163	0.054	0.048	0.058	-0.01	6,079
	(0.415)	(0.032)*	(0.04)	(0.039)	(0.047)	

Notes: Standard errors clustered on section. We drop a single outlying observation from Konjo Njeigor for the week of August 24, 2014, which is 25 times larger than any other weekly total from that section and 5 times larger than any other observation in the full time-series. Konjo Njeigor is a CM section, so removing this observation only depresses the pooled and CM treatment effects. Treatment effects estimated using OLS including matching-triplet and week fixed effects. Significance: * $p < 0.10$, ** $p < 0.05$, and *** $p < 0.01$.

E.10 Cross-sectional Results for Reported Cases

Table E.10: Effect on Reported Cases (Cross-Sectional)

	Control Mean	Pooled	CM	NFA	Difference	N
Ebola Cases						
Total	9.537	5.868	6.925	5.047	1.878	160
	(12.462)	(4.086)	(4.992)	(4.661)	(5.037)	
Confirmed	0.389	2.012	2.909	1.315	1.594	160
	(1.235)	(1.197)*	(1.453)**	(1.357)	(1.467)	
Negative	8.093	3.383	2.689	3.922	-1.233	160
	(10.617)	(2.916)	(3.563)	(3.327)	(3.595)	
Log(Ebola Cases + 1)						
Total	1.798	0.294	0.417	0.198	0.219	160
	(1.11)	(0.214)	(0.26)	(0.243)	(0.262)	
Confirmed	0.171	0.38	0.413	0.354	0.059	160
	(0.456)	(0.141)***	(0.172)**	(0.161)**	(0.174)	
Negative	1.674	0.227	0.348	0.133	0.215	160
	(1.07)	(0.202)	(0.245)	(0.229)	(0.247)	
Linear Probability Model						
Total	0.833	0.072	0.091	0.057	0.034	160
	(0.376)	(0.062)	(0.076)	(0.071)	(0.076)	
Confirmed	0.148	0.251	0.234	0.265	-0.03	160
	(0.359)	(0.082)***	(0.101)**	(0.094)***	(0.102)	
Negative	0.833	0.048	0.096	0.011	0.086	160
	(0.376)	(0.067)	(0.081)	(0.076)	(0.082)	
IHS(Ebola Cases)						
Total	2.242	0.335	0.484	0.22	0.264	160
	(1.328)	(0.25)	(0.304)	(0.284)	(0.307)	
Confirmed	0.221	0.474	0.505	0.45	0.055	160
	(0.587)	(0.173)***	(0.212)**	(0.198)**	(0.214)	
Negative	2.097	0.264	0.424	0.14	0.285	160
	(1.291)	(0.239)	(0.291)	(0.271)	(0.293)	
Poisson						
Total	9.537	0.469	0.552	0.407	0.144	160
	(12.462)	(0.056)***	(0.066)***	(0.062)***	(0.062)**	
Confirmed	0.389	1.669	2.008	1.369	0.639	160
	(1.235)	(0.231)***	(0.244)***	(0.245)***	(0.152)***	
Negative	8.093	0.342	0.276	0.389	-0.113	160
	(10.617)	(0.062)***	(0.075)***	(0.069)***	(0.073)	

Notes: Treatment effects estimated using OLS including matching-triplet fixed effects. Significance: * p < 0.10, ** p < 0.05, and *** p < 0.01

Figure E.7: Cross-sectional Differences in Confirmed Cases

(a) Empirical Cumulative Distribution

(b) Unadjusted Means
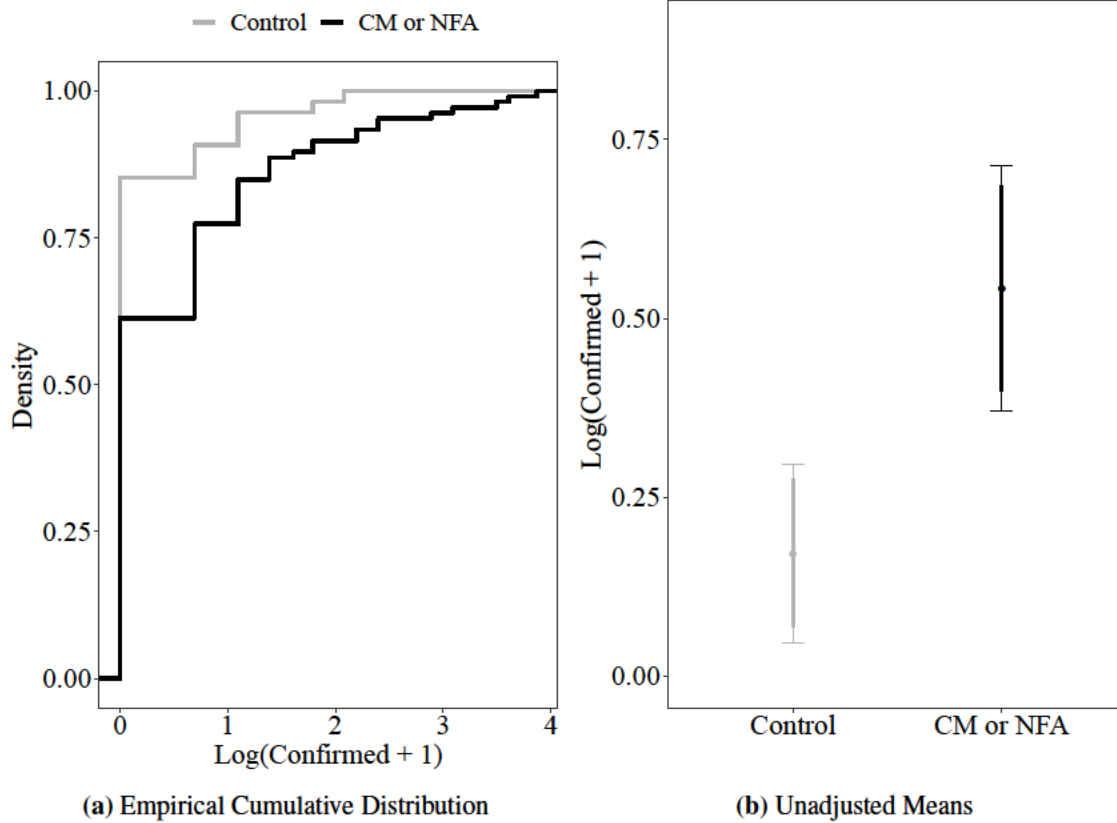
Figure E.7(a): empirical CDF of reported, confirmed cases (logged) for control (grey) and treated (black) sections. Figure E.7(b): average number of reported, confirmed cases (logged) for control and treated sections. These means do not account for the blocking.

E.11 Dose-response Models

Table E.11: Dose-Response with All Sections in Study Area

	Total		Total per Clinic		Total per 1k	
	(1)	(2)	(3)	(4)	(5)	(6)
Proportion of Clinics Treated	0.173**	0.159*	0.165**	0.169**	0.052*	0.053*
	(0.084)	(0.084)	(0.084)	(0.082)	(0.031)	(0.031)
Population (1000s)				0.059***		
				(0.016)		
Number of Clinics						0.074***
						(0.015)
	Confirmed		Confirmed per Clinic		Confirmed per 1k	
	(1)	(2)	(3)	(4)	(5)	(6)
Proportion of Clinics Treated	0.059**	0.059**	0.059**	0.059**	0.015**	0.015**
	(0.024)	(0.024)	(0.024)	(0.024)	(0.006)	(0.006)
Population (1000s)				0.007		
				(0.005)		
Number of Clinics						0.008**
						(0.003)
Ebola Sample	✓					
Full Sample		✓	✓	✓	✓	✓
Sections	160	205	205	205	205	205
Observations	5,440	6,970	6,970	6,970	6,970	6,970

Notes: Standard errors clustered on section. Treatment effects estimated using OLS including matching-triplet and week fixed effects. Significance: * $p < 0.10$, ** $p < 0.05$, and *** $p < 0.01$.

E.12 Placebo Test with Nearest Neighboring Out-of-sample Sections

We calculate the distances between the centroid of a section in the sample and all out-of-sample sections. We then use the minimum distance to identify the nearest neighbor.

Table E.12: Placebo: Reported Cases using Nearest Neighboring Section

	Control Mean	Pooled	CM	NFA	Difference	N
Ebola Cases						
Total	0.222	0.057	0.043	0.069	-0.026	5,440
	(0.957)	(0.063)	(0.076)	(0.072)	(0.078)	
Confirmed	0.029	0.01	0.001	0.016	-0.015	5,440
	(0.371)	(0.023)	(0.03)	(0.027)	(0.033)	
Negative	0.169	0.042	0.033	0.049	-0.015	5,440
	(0.653)	(0.04)	(0.047)	(0.046)	(0.047)	
IHS(Ebola Cases)						
Total	0.153	0.041	0.033	0.048	-0.015	5,440
	(0.419)	(0.031)	(0.036)	(0.035)	(0.036)	
Confirmed	0.018	0.008	0.004	0.012	-0.009	5,440
	(0.164)	(0.012)	(0.016)	(0.014)	(0.017)	
Negative	0.125	0.032	0.024	0.038	-0.013	5,440
	(0.37)	(0.025)	(0.029)	(0.028)	(0.028)	

Notes: Standard errors clustered on section. Treatment effects estimated using OLS including matching-triplet and week fixed effects. Significance: * p < 0.10, ** p < 0.05, and *** p < 0.01. The three major cities in our study districts (Bo Town, Kenema Town, and Makeni Town) are excluded as potential nearest neighbors.

E.13 Baseline Balance in Ebola Sample

Table E.13: Baseline Balance (Ebola sub-sample)

	(1) Control Mean	(2) CM	(3) NFA	(4) Difference	(5) N
Village characteristics					
Motorable road	0.891	0.108	0.045	0.063	318
	(0.313)	(0.059)*	(0.050)	(0.055)	
Mobile phone coverage	0.812	0.078	0.174	−0.096	318
	(0.392)	(0.076)	(0.058)***	(0.065)	
Distance to the closest clinic	1.362	0.338	0.909	−0.571	318
	(2.217)	(0.532)	(0.905)	(0.505)	
Travel cost to closest clinic	94.225	−29.666	−46.167	16.501	317
	(869.677)	(132.127)	(124.628)	(102.778)	
Household characteristics and questions to household head					
Household size	3.369	−0.038	0.016	−0.053	6225
	(2.979)	(0.032)	(0.031)	(0.032)	
Number of illness or injury cases per household	0.054	−0.021	−0.013	−0.009	6225
	(0.237)	(0.009)**	(0.009)	(0.009)	
Birth in household last year	0.157	−0.024	0.011	−0.035	1582
	(0.363)	(0.020)	(0.019)	(0.019)*	
Child under 2 in household	0.230	−0.025	0.001	−0.026	1581
	(0.421)	(0.027)	(0.024)	(0.025)	
Prominent village member in household	0.042	−0.019	−0.016	−0.003	1574
	(0.200)	(0.014)	(0.013)	(0.014)	
Believes doctor's advice	0.995	−0.001	−0.005	0.004	1452
	(0.072)	(0.007)	(0.007)	(0.006)	
Health care fees unaffordable	2.307	−0.005	0.070	−0.075	1521
	(0.784)	(0.069)	(0.068)	(0.074)	
Trust in the community	1.856	−0.129	−0.054	−0.075	1581
	(0.663)	(0.063)**	(0.057)	(0.070)	
Community cohesion	2.420	0.017	0.025	−0.009	1577
	(0.610)	(0.045)	(0.048)	(0.051)	
Believe VHC members represent your interest	2.743	−0.117	−0.174	0.057	704
	(1.061)	(0.122)	(0.123)	(0.113)	
The VHC can be trusted	2.453	−0.121	0.026	−0.148	843
	(0.967)	(0.120)	(0.107)	(0.122)	
Individual characteristics					
Muslim	0.854	0.000	0.001	−0.001	7128
	(0.353)	(0.040)	(0.034)	(0.045)	
Mende (Ethnicity)	0.418	−0.029	0.011	−0.040	7127
	(0.493)	(0.017)*	(0.016)	(0.018)**	
Temne (Ethnicity)	0.348	0.104	0.121	−0.018	7127
	(0.476)	(0.056)*	(0.048)**	(0.056)	
Highest level of education	2.920	0.054	0.780	−0.726	7106
	(4.203)	(0.176)	(0.161)***	(0.180)***	

Notes: This table illustrates the baseline balance across the three treatment arms (control, community monitoring and non-financial awards). Column (1) shows the characteristics of the control group at baseline (mean and standard deviation). Columns (2) and (3) indicate the regression coefficients and standard errors of the CM and NFA treatment arms compared to the control group. Column (4) compares the two treatment arms. Significance: * is significant at the 10% level, ** is significant at the 5% level and *** is significant at the 1% level.

E.14 Results for Pre-specified Families in Ebola Sample

Table E.14: Pre-specified Families in Ebola Sample

	(1) Control Mean	(2) Pooled	(3) CM	(4) NFA	(5) N
General utilization	0.000	0.074	0.128	0.036	2857
	(1.000)	(0.040)*	(0.049)***	(0.042)	
		[0.188]	[0.110]	[0.397]	
Maternal utilization index	0.000	0.058	0.232	−0.055	595
	(1.000)	(0.086)	(0.112)**	(0.097)	
		[0.379]	[0.143]	[0.488]	
Health outcomes index	0.000	0.034	0.085	−0.006	3183
	(1.000)	(0.062)	(0.069)	(0.074)	
		[0.379]	[0.283]	[0.886]	
Satisfaction index	0.000	0.108	0.106	0.109	3183
	(1.000)	(0.055)*	(0.068)	(0.061)*	
		[0.188]	[0.235]	[0.144]	
Health service delivery index	0.000	−0.056	−0.033	−0.064	1819
	(1.000)	(0.066)	(0.088)	(0.070)	
		[0.329]	[0.459]	[0.397]	
Clinic quality index	0.000	0.311	−0.023	0.565	160
	(1.000)	(0.236)	(0.301)	(0.264)**	
		[0.205]	[0.459]	[0.144]	
CDPE index	0.000	0.264	0.232	0.288	320
	(1.000)	(0.108)**	(0.155)	(0.132)**	
		[0.188]	[0.235]	[0.144]	
Contributions to clinic index	0.000	0.173	0.292	0.080	320
	(1.000)	(0.127)	(0.142)**	(0.148)	
		[0.205]	[0.143]	[0.488]	
Water and sanitation index	0.000	0.167	0.137	0.190	3183
	(1.000)	(0.086)*	(0.101)	(0.096)*	
		[0.188]	[0.264]	[0.144]	
Economic outcomes index	0.000	0.151	0.065	0.219	3183
	(1.000)	(0.087)*	(0.106)	(0.105)**	
		[0.188]	[0.425]	[0.144]	

Notes: Treatment effects are estimated using Missing-Indicator ANCOVA, controlling for the community-level average of the outcome family index at baseline and matching-triplet fixed effects. Robust standard errors, clustered by clinic, are shown in parentheses. Multiple-inference corrected q-values that adjust for the false discovery rate within treatment arm are shown in square brackets. Significance: * is significant at the 10% level, ** is significant at the 5% level and *** is significant at the 1% level.

E.15   Perceived Quality of Care

Table E.15: Perceived quality of care

	Control Mean	Pooled	CM	NFA	N
General utilization	0.972	0.026	0.048	0.010	2857
	(0.398)	(0.015)*	(0.018)***	(0.016)	
Satisfaction with public health workers	3.301	0.103	0.100	0.105	3149
	(0.779)	(0.036)***	(0.048)**	(0.040)**	
Relative effectiveness of western medicine	−0.378	0.060	0.070	0.053	3183
	(0.205)	(0.025)**	(0.033)**	(0.027)*	

Notes: Robust standard errors are in parentheses. Standard errors are clustered by clinic. Treatment effects are estimated using Missing Indicator ANCOVA, controlling for the outcome family index at baseline at the community level, and including matching-triplet fixed effects. Significance: * is significant at the 10% level, ** is significant at the 5% level and *** is significant at the 1% level.

E.16 Surveillance

The WHO defines contact tracing as: "identification and follow-up of persons who may have come into contact with a person infected with the Ebola virus."[A6]

Table E.16: Contact Tracing among Confirmed Patients by Treatment

Treatment	Total Traced	Proportion Traced	Proportion among Contacts	
			Family	Outside
Control	17	0.59	0.50	0.50
CM	55	0.22	0.61	0.39
NFA	28	0.24	0.57	0.43

Notes: Total Traced (column 2) counts the number of cases subject to contact tracing across the treatment arms. Proportion traced then divides this number by the total number of confirmed cases. In the final two columns, we restrict attention to cases subject to contact tracing and compute the proportion of contacts from the patients' family or outside their family. Family here includes individuals within the nuclear family, e.g., parents, children, siblings.

Table E.17: Effect on Surveillance Proxies

	Control Mean	Pooled	CM	NFA	Difference	N
Pr(Lab Test)	0.926	0.012	0.018	0.008	0.01	144
	(0.099)	(0.016)	(0.02)	(0.018)	(0.02)	
Delay: Report - Lab	5.029	-2.874	-3.072	-2.72	-0.352	160
	(16.915)	(1.978)	(2.419)	(2.259)	(2.441)	
Log(Case Workers + 1)	1.672	0.125	0.134	0.117	0.016	160
	(0.952)	(0.18)	(0.221)	(0.206)	(0.223)	

Notes: Treatment effects estimated using OLS including matching-triplet fixed effects. Significance: * $p < 0.10$, ** $p < 0.05$, and *** $p < 0.01$

[A6] https://www.who.int/csr/resources/publications/ebola/contact-tracing-guidelines/en/

E.17 Ebola-specific Balance Tests

Table E.18: Balance: Specialized Ebola Facilities

	Control Mean	Pooled	CM	NFA	N
NERC					
Total	0.056	-0.03	-0.054	-0.011	160
	(0.231)	(0.042)	(0.051)	(0.048)	
EHC	0.019	0.006	-0.016	0.023	160
	(0.136)	(0.033)	(0.04)	(0.037)	
CCC	0.019	-0.018	-0.027	-0.011	160
	(0.136)	(0.021)	(0.026)	(0.024)	
Beds	0.463	-0.18	-0.66	0.194	160
	(2.044)	(0.493)	(0.596)	(0.556)	
UNMEER					
Total	0.093	0.03	0.024	0.035	160
	(0.293)	(0.065)	(0.079)	(0.074)	
EHC	0.019	-0.012	-0.013	-0.011	160
	(0.136)	(0.029)	(0.035)	(0.033)	
CCC	0.056	0.06	0.048	0.069	160
	(0.231)	(0.056)	(0.068)	(0.064)	
Beds	0.759	0.347	0.351	0.345	160
	(2.495)	(0.623)	(0.762)	(0.712)	

Notes: Differences estimated using OLS including matching-triplet fixed effects. Significance: * $p < 0.10$, ** $p < 0.05$, and *** $p < 0.01$

Table E.19: Balance: Minimum Distance to Specialized Ebola Facilities

	Control Mean	Pooled	CM	NFA	N
NERC					
ETU	33.986	0.588	-1.302	2.056	160
	(17.422)	(3.795)	(4.626)	(4.319)	
EHC	20.715	-2.034	-0.315	-3.369	160
	(11.479)	(2.115)	(2.563)	(2.394)	
CCC	49.794	-5.353	-3.257	-6.98	160
	(31.244)	(4.09)	(4.984)	(4.654)	
UNMEER					
ETU	33.57	0.82	-1.156	2.354	160
	(17.495)	(3.777)	(4.602)	(4.297)	
EHC	20.586	-2.361	-1.105	-3.337	160
	(10.445)	(2.011)	(2.446)	(2.284)	
CCC	54.163	-7.277	-5.159	-8.921	160
	(43.64)	(3.819)*	(4.651)	(4.343)**	

Notes: Differences estimated using OLS including matching-triplet fixed effects. Significance: * $p < 0.10$, ** $p < 0.05$, and *** $p < 0.01$

Table E.20: Balance: Proxies for Exposure

	Control Mean	Pooled	CM	NFA	N
Dist(Patient Zero in Guinea)	196.354	10.93	8.158	13.083	160
	(42.021)	(5.079)**	(6.186)	(5.777)**	
Dist(Patient Zero in SL)	91.104	11.627	7.758	14.632	160
	(62.208)	(6.441)*	(7.838)	(7.319)**	
Primary Road Density	0.007	0.008	-0.012	0.023	160
	(0.029)	(0.015)	(0.018)	(0.017)	
Secondary Road Density	0.019	0.011	0.013	0.01	160
	(0.053)	(0.011)	(0.014)	(0.013)	
Tertiary Road Density	0.084	-0.009	0.01	-0.024	160
	(0.158)	(0.024)	(0.029)	(0.027)	

Notes: Differences estimated using OLS including matching-triplet fixed effects. Significance: * $p < 0.10$, ** $p < 0.05$, and *** $p < 0.01$

E.18 Spillovers

Table E.21: Spillovers from Bordering Sections

	Total			IHS(Total)		
	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Pooled	0.173**	0.360**	0.330***	0.083*	0.179**	0.184***
	(0.084)	(0.177)	(0.113)	(0.043)	(0.089)	(0.059)
Pooled × Bordering Controls		-0.134			-0.070	
		(0.116)			(0.060)	
Bordering Controls		0.105			0.052	
		(0.083)			(0.044)	
Pooled × Bordering Controls Pop. (1000s)			-0.019			-0.012
			(0.021)			(0.011)
Bordering Control Pop. (1000s)			0.050**			0.032***
			(0.020)			(0.010)
Week	34	34	34	34	34	34
Block	81	81	81	81	81	81
N	5,440	5,440	5,440	5,440	5,440	5,440

Notes: Standard errors clustered on section. Treatment effects estimated using OLS including matching-triplet and week fixed effects. Significance: * $p < 0.10$, ** $p < 0.05$, and *** $p < 0.01$.

Table E.22: Spillovers from Clinic Proximity

	Total		IHS(Total)	
	Model 1	Model 2	Model 3	Model 4
Pooled	0.173**	0.809**	0.083*	0.473**
	(0.084)	(0.385)	(0.043)	(0.207)
Pooled × Proximity to Control		-0.028*		-0.017*
		(0.017)		(0.009)
Proximity to Control		0.010		0.007
		(0.011)		(0.006)
Week	34	34	34	34
Block	81	81	81	81
N	5,440	5,440	5,440	5,440

Notes: Standard errors clustered on section. Treatment effects estimated using OLS including matching-triplet and week fixed effects. To compute proximity, we measure the distance (in kilometers) to the nearest control clinic in the full sample and then reverse the scale of the variable by subtracting off the maximum and multiplying by minus one. Significance: *p < 0.10, ** p < 0.05, and *** p < 0.01.

Table E.23: Spillovers through Road Network

	Total		IHS(Total)	
	Model 1	Model 2	Model 3	Model 4
Pooled	0.173**	0.177	0.083*	0.099*
	(0.084)	(0.114)	(0.043)	(0.056)
Pooled × Connected Controls		0.026		0.011
		(0.025)		(0.013)
Connected Controls		0.001**		0.001***
		(0.000)		(0.000)
Week	34	34	34	34
Block	81	81	81	81
N	5,440	5,440	5,440	5,440

Notes: Standard errors clustered on section. Treatment effects estimated using OLS including matching-triplet and week fixed effects. Significance: * $p < 0.10$, ** $p < 0.05$, and *** $p < 0.01$.

Table E.24: Spillovers from Proximate Coethnic Sections

	Total		IHS(Total)	
	Model 1	Model 2	Model 3	Model 4
Pooled	0.173**	0.243**	0.083*	0.123**
	(0.084)	(0.117)	(0.043)	(0.058)
Pooled × Co-ethnic Controls w/in 10 km		-0.170		-0.090
		(0.111)		(0.061)
Co-ethnic Controls w/in 10 km		0.120		0.056
		(0.083)		(0.048)
Week	34	34	34	34
Block	81	81	81	81
N	5,440	5,440	5,440	5,440

Notes: Standard errors clustered on section. Treatment effects estimated using OLS including matching-triplet and week fixed effects. Significance: *p < 0.10, ** p < 0.05, and *** p < 0.01.

Figure E.8: Roads Intersecting Control Sections

Black Paths: Roads and paths intersecting control sections
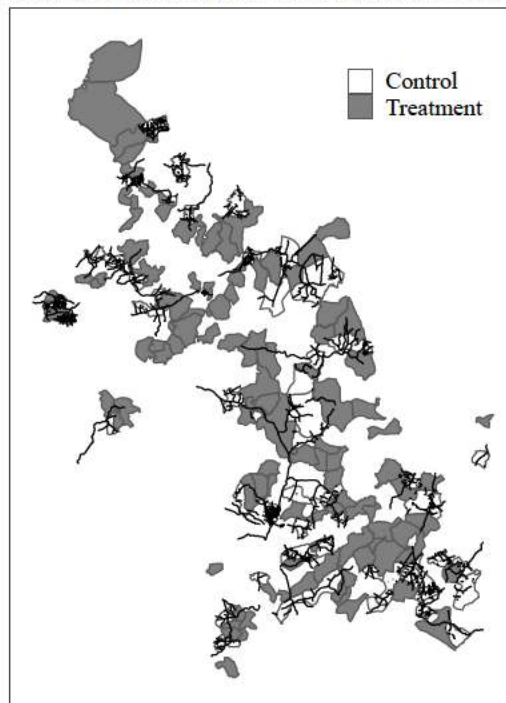
E.19 Ratio of Confirmed and Total Cases

Figure E.9: Ratio of Confirmed to Total Cases

(a) Ratio Confirmed to Total

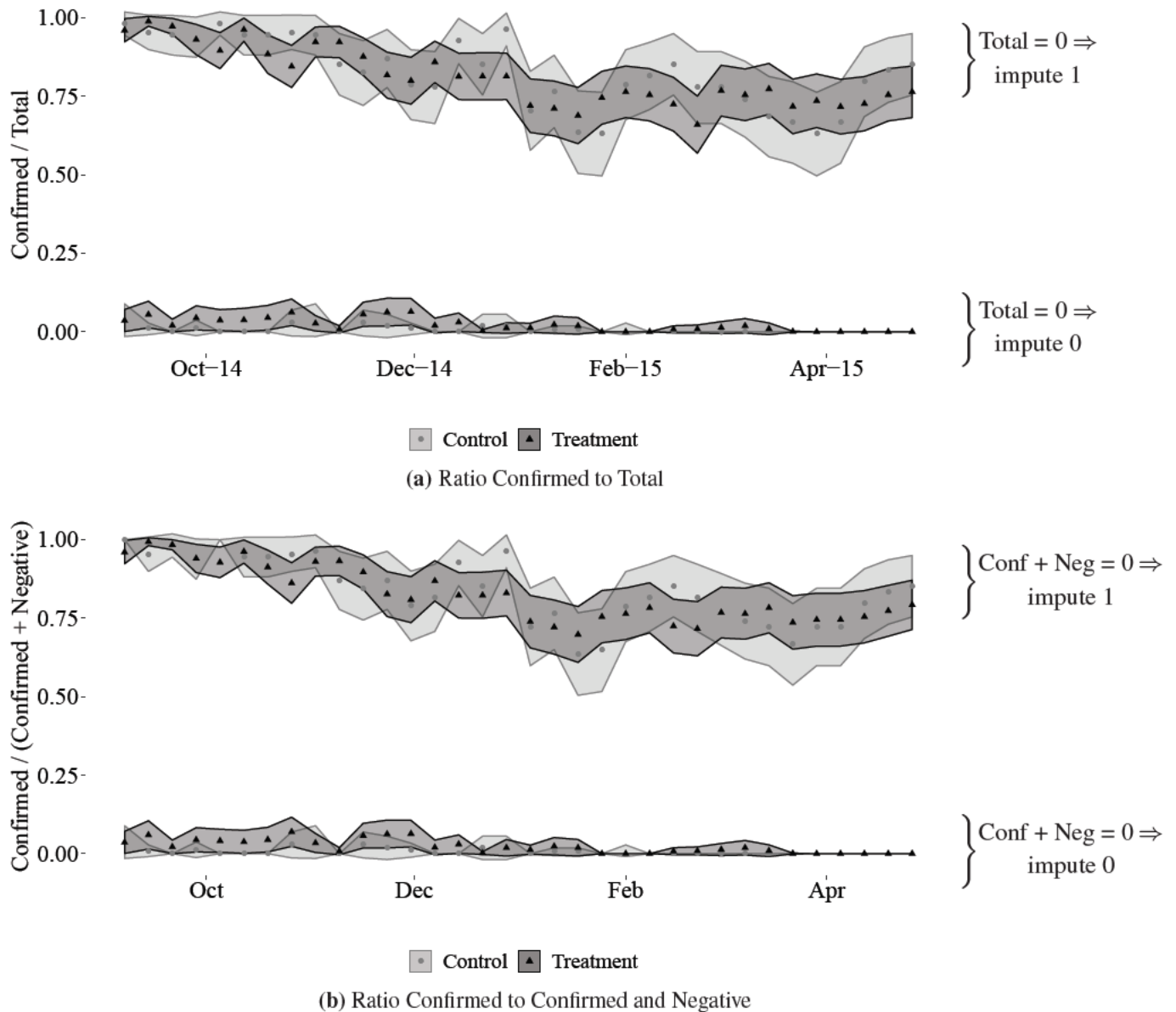(b) Ratio Confirmed to Confirmed and Negative

Figure E.9(a) computes the ratio of confirmed to total cases for each section-week and then average across treatment and control. If there are no cases in a section-week, the ratio is undefined. The ribbons at the top of the plot display the averages when we impute 1 for those undefined observations; the ribbons at the bottom display the averages when we instead impute 0. Figure E.9(b) computes the ratio of confirmed to confirmed plus negative cases. If there the sum of confirmed and negative cases is zero in a section-week, the ratio is undefined. The ribbons at the top of the plot display the averages when we impute 1 for those undefined observations; the ribbons at the bottom display the averages when we instead impute 0. In both figures, the shaded areas connect the 95% confidence intervals around these proportions.

E.20 Bounding Exercise: Unintended Increase

Data on Ebola incidence in Sierra Leone is incomplete. As such, we cannot directly rule out an increase in exposure by comparing the total number of cases in treatment and control areas. To be a confirmed case in the available Ebola data, an individual must be infected with Ebola and known to health workers through self-reporting or surveillance. CM and NFA could theoretically affect case counts by unintentionally increasing either exposure rates, reporting propensities, or both. We use our empirical results and a simple model to clarify what must be assumed to attribute our results to changes in exposure.

Sequence and Information

Each individual i observes whether they are symptomatic, $s \in \{0,1\}$. The CDC lists the following as Ebola symptoms: fever, severe headache, muscle pain, weakness or fatigue, diarrhea, vomiting abdominal pain, or unexplained hemorrhage. They also observe the treatment status of their local health facility, $T \in \{0,1\}$.

$\quad i$ knows that $\Pr[s = 1 \mid I = 1] = 1$: if you have Ebola, you will show symptoms. They also know that $\Pr[s = 1 \mid I = 0] = p \in (0,1)$, i.e., that symptoms like fevers and diarrhea happen to those that are not infected.[A7] The infection rates within control and treatment communities, $e_T = \mathbb{E}(I \mid T)$, are also common knowledge.

$\quad i$ must to decide whether to report their symptoms and be tested, $R \in \{0,1\}$. They cannot, however, condition this decision on their actual infection status, because this is not known to i prior to testing.

Notation

- Reporting among Symptomatic in Control: $\Pr(R \mid s = 1, T = 0) = h \in [0,1]$

- Reporting among Symptomatic in Treatment: $\Pr(R \mid s = 1, T = 1) = \min\{h\tau_h, 1\}$ where $\tau_h \in \mathbb{R}^1_+$

- Reporting among Asymptomatic in Control: $\Pr(R \mid s = 0, T = 0) = l \in [0,1]$

- Reporting among Asymptomatic in Treatment: $\Pr(R \mid s = 0, T = 1) = \min\{l\tau_l, 1\}$ where $\tau_l \in \mathbb{R}^1_+$

$\quad$ We assume $l \leq h$ (i.e., individuals with symptoms are more likely to report than those without). To minimize terms, we define $d = l/h$. This is the ratio of reporting probabilities of asymptomatic to symptomatic individuals in control areas. $d = 0.5$, for example, implies that symptomatic individuals in control areas are twice as likely to report as those displaying no symptoms.

[A7]This is likely a considerable proportion of Sierra Leone's population: the WHO, for example, cites a 2011 assessment which found that 24 percent of children under 5 had malaria in the *two weeks* prior to the survey. Over months, the probability of flu-like symptoms due to illnesses unrelated to Ebola is quite likely.

Logic

We estimate the percentage difference in confirmed cases between treatment and control β where:

$$\mathbb{E}[R * I \mid T = 1] = \beta \; \mathbb{E}[R * I \mid T = 0] \qquad (4)$$
$$\beta = \frac{e_1}{e_0} \tau_h = \mathscr{E} \tau_h$$

where $\mathscr{E}$ is the effect of the treatment on exposure to Ebola. This estimate could confound the effect of the treatment on exposure $\mathscr{E}$ and reporting τ_h by symptomatic individuals.

In the cross-sectional results, we estimate $\widehat{\beta} \approx 1.4$. If we make the extreme assumption that treatment has no impact on the reporting decisions of symptomatic individuals ($\tau_h \to 1$), then β reflects the different rates of exposure in treatment and control areas. Conversely, as $\tau_h \to \beta$, the possible treatment effect on exposure attenuates to zero ($\mathscr{E} \to 1$).

Second, we find that the ratio of confirmed to total cases does not differ with treatment status:

$$\frac{\mathbb{E}[R * I \mid T = 1]}{\mathbb{E}[R \mid T = 1]} = \frac{\mathbb{E}[R * I \mid T = 0]}{\mathbb{E}[R \mid T = 0]}$$

Rearranging equation (1) and substituting,

$$\beta = \frac{\mathbb{E}[R \mid T = 1]}{\mathbb{E}[R \mid T = 0]}$$

This implies

$$\tau_l = \tau_h \left(\frac{\mathscr{E} d (1 - e_0)(1 - p) + p(\mathscr{E} - 1)}{d(1 - e_1)(1 - p)} \right)$$
$$\text{and}$$
$$\mathscr{E} = \beta / \tau_h$$

If treatment increased exposure, then it must have also increased reporting among asymptomatic individuals, such that the ratio of confirmed to total cases is not elevated in treated areas.

Using these two equations, we vary the parameters $\{\tau_h, p, d\}$ over plausible ranges and compute the implied increases in exposure ($\mathscr{E}$) and reporting among asymptomatic individuals (τ_l). We set $\beta = \widehat{\beta} \approx 1.4$ and $e_1 = 0.01$.[A8] This exercise clarifies what we must be true for the treatment to increase communities' exposure to Ebola ($\mathscr{E} > 1$) and still produce our empirical results:

- There is some pathway whereby treatment increased exposure.

- $\tau_h < \beta$. As $\tau_h \to \beta$, the potential positive effect on exposure attenuates to zero (i.e., $\mathscr{E} \to 1$).

- $\tau_h < \tau_l$. Treatment must have had a *larger* effect on reporting among those *without* symptoms.

[A8]The exact value of e_1 is not consequential at low values of e_1. We do not need to set e_0, as this is equal to $\beta \tau_h / e_1$.

- This differential effect (τ_l/τ_h) must be larger when d is smaller or p is larger. If baseline reporting is much lower among asymptomatic individuals and/or Ebola symptoms are common among uninfected individuals, then τ_l/τ_h must be large.

Numerical Examples

Suppose that individuals with no symptoms report 25 percent as often as those with symptoms ($d = 0.25$) and that 25 percent of individuals display flu-like symptoms over the course of several months even when uninfected ($p = 0.25$). If treatment has no effect on reporting among symptomatic individuals ($\tau_h = 1$), then the 40 percent increase in exposure would have to be accompanied by roughly two times as much reporting by individuals with no symptoms ($\tau_l = 1.94$). If treatment led to a 20 percent increase in reporting among those with symptoms ($\tau_h = 1.2$), then exposure can only increase by 16 percent, and τ_l must reach 1.67.

Table E.25: Implied $\mathscr{E}$ and τ_l

p	d	τ_h	τ_l	$\mathscr{E}$
0.25	0.25	1.00	1.94	1.40
0.50	0.25	1.00	3.02	1.40
0.25	0.50	1.00	1.67	1.40
0.50	0.50	1.00	2.21	1.40
0.25	0.25	1.20	1.67	1.17
0.50	0.25	1.20	2.21	1.17
0.25	0.50	1.20	1.54	1.17
0.50	0.50	1.20	1.81	1.17

For reasonable choices of p and d, we find these scenarios unlikely. First, in our NFA treatment, there is no plausible pathway whereby treatment increased exposure. Even in CM areas, the last planned community meeting took place a year prior to the Ebola outbreak.

Second, these imply large treatment effects and high relative rates of reporting among individuals with no symptoms to report. One cannot preemptively test for Ebola — the virus can only be detected days after symptoms begin. There is no reason for asymptomatic individuals to report, and widespread fear that deterred the use of health facilities even among those in desperate need of medical care. Elston et al. (2016, 675) report reductions in hospital attendance during the Ebola crisis in Sierra Leone, including significantly lower numbers of "women admitted during labor, urgent paediatric hospital admissions including children hospitalized with malaria and outpatient consultations."

Third, ceiling effects are unlikely. One might be concerned that h is close to 1. In that case, there is less room for treatment to affect reporting among those displaying symptoms, as $h\tau_h \leq 1$. However, the CDC forecasts used an underreporting factor of 2.5 for Sierra Leone and Liberia based on expert opinions.[A9] (This would correspond to $h = 0.4$ in our terms.) This implies that τ_h could be as large as 2.5 before hitting any ceiling effects. We can rule out any increase in exposure when $\tau_h \geq 1.4$. Qualitative evidence stresses underreporting as a major concern during Sierra Leone's Ebola crisis. This implies that many symptomatic

---

[A9] https://www.cdc.gov/mmwr/pdf/other/su6303.pdf

individuals were failing to seek care and, thus, might have changed their decision as a consequence of treatment.

The data do not allow us to rule out an increase in exposure. However, to reconcile this explanation with our pattern of results requires behavioral responses among asymptomatic individuals that we find difficult to believe.

E.21 Controlling for Unbalanced Baseline Variables

Covariates included (all measured as proportions of households): (1) beyond primary education, (2) Mende, (3) ill in last month, and (4) own mobile phone.

Table E.26: Effects on Reported Cases Controlling for Unbalanced Baseline Variables

	Control Mean	Pooled	CM	NFA	Difference	N
Ebola Cases						
Total	0.281	0.191	0.153	0.23	-0.077	5,440
	(0.727)	(0.086)**	(0.103)	(0.101)**	(0.112)	
Confirmed	0.011	0.068	0.07	0.067	0.002	5,440
	(0.129)	(0.027)**	(0.031)**	(0.029)**	(0.027)	
Negative	0.238	0.105	0.056	0.153	-0.098	5,440
	(0.648)	(0.063)*	(0.076)	(0.078)**	(0.089)	
IHS(Ebola Cases)						
Total	0.206	0.096	0.082	0.11	-0.029	5,440
	(0.47)	(0.043)**	(0.053)	(0.051)**	(0.059)	
Confirmed	0.009	0.033	0.029	0.037	-0.007	5,440
	(0.1)	(0.011)***	(0.013)**	(0.013)***	(0.013)	
Negative	0.179	0.066	0.047	0.086	-0.038	5,440
	(0.433)	(0.036)*	(0.045)	(0.043)**	(0.051)	

Notes: Standard errors clustered on section. Treatment effects estimated using OLS including matching-triplet and week fixed effects. Covariates included (all measured as proportions of households): (1) beyond primary education, (2) Mende, (3) ill in last month, and (4) own mobile phone. Significance: * $p < 0.10$, ** $p < 0.05$, and *** $p < 0.01$.

E.22   Geo-coding Procedure

The VHF data includes information on individuals' residences, including their district, chiefdom, and village or parish. We use this information to place observations within sections. Our geo-location protocol involves several steps. First, a human coder inspected and cleaned all district and chiefdom names that did not exactly match the conventional spelling. Of 85,410 entries in the case data, we can code the chiefdom of residence for 97% of observations.

Second, we employ fuzzy string matching to match the available village or parish names to gazetteer files of placenames from Sierra Leone. Fortunately, in the chiefdoms that include our sample, only 14 confirmed, suspected, or probable Ebola cases do not include village or parish information.[A10] We employ the gazetteer file from Open Street Map (www.openstreetmap.org/), which includes 9,975 entries, ranging from hamlets to cities. We prefer this list to the 2004 census data from Sierra Leone, which only provides names for around 5,000 localities. Moreover, during the Ebola epidemic, Open Street Map mounted a humanitarian effort aimed at updating and verifying information on the locations of villages and roads in Sierra Leone.[A11]

Ten sample entries from OSM gazetteer file:

```
       osm_id           name            coordinates
 1  27565056  Freetown     (-13.26802 8.479002)
 2  314001434 Bo           (-11.73665 7.962065)
 3  314005602 Kenema       (-11.18639 7.885936)
 4  314007819 Koidu        (-10.97163 8.642281)
 5  320058940 Kambia       (-12.91934 9.125073)
 6  320060481 Kamakwie     (-12.24125 9.496301)
 7  320060535 Pujehun      (-11.72124 7.356632)
 8  320060540 Zimmi        (-11.31032 7.312338)
 9  370327499 Goderich     (-13.28887 8.432966)
10  370495828 Murray Town (-13.26534 8.491613)
```

Fuzzy string matching calculates the string distance between each village or parish name in the VHF data and each placename in the gazetteer file that falls within the exact same district and chiefdom.[A12] An exact match returns a distance of zero; "FREE TOWN" and "FREETOWN," for example, would return a distance of 1. We do not match any entries with a string distance that exceeds 2.

While the geo-coding process introduces measurement error, we expect this is uncorrelated with treatment and, thus, only going to attenuate our estimates. To bolster this assumption, we look at whether placenames in the gazetteer file tend to be more numerous or longer in treated versus control sections. We see no indication that treated sections have significantly more or shorter placenames; moreover, the placenames are not more likely to contain a space between words (see Table E.27).

[A10] Of all entries in the case data that fall within the chiefdoms the include our sample, only 0.07 percent are missing an entry for village or parish of residence.

[A11] http://wiki.openstreetmap.org/wiki/2014_West_Africa_Ebola_Response

[A12] We use optimal string alignment distance, a variant of the Levenshtein distance, which is commonly employed in geo-coding algorithms.

Table E.27: Balance: Placenames for Geocoding

	Control Mean	Pooled	CM	NFA	N
Number of Places	8.056	1.096	0.721	1.387	160
	(6.456)	(1.303)	(1.592)	(1.487)	
Number of Placenames	7.222	0.856	0.651	1.016	160
	(5.709)	(1.304)	(1.594)	(1.489)	
Average Length of Placenames	6.443	0.021	0.085	-0.029	160
	(1.945)	(0.3)	(0.367)	(0.343)	
Proportion of Placenames with Whitespace	0.033	-0.004	-0.008	0	160
	(0.114)	(0.015)	(0.018)	(0.017)	

Notes: Differences estimated using OLS including matching-triplet fixed effects. Significance: * $p < 0.10$, ** $p < 0.05$, and *** $p < 0.01$

E.23 Calculating Reduction in R_0

R_0 is the reproduction rate of a disease: the average number of secondary cases generated by the average infectious individual. To calculate the implied reduction in R_0 due to our treatments we follow the approach of Pronyk et al. (2016), which the authors detail in their online appendix.[A13] R_0 is calculated by multiplying the disease transmission rate by the average duration of infectiousness, $D \geq 0$. The duration of infectiousness is time during which an infected patient can spread disease.

We adopt Pronyk et al.'s (2016) assumption that transmission rates do not change following public health interventions (in their case the construction of Community Care Centers). Conditional on a infected individual and susceptible individual coming into contact, the likelihood that Ebola is transmitted between the two is unaffected by treatment. Granting this assumption, treatment can affect R_0 by changing $D(T)$, which is calculated as follows:

$$D(T) = t(T)r(T) + 10[1 - r(T)]$$

where $t(T)$ is the time between symptom onset and isolation among individuals who are isolated; $r(T)$, the proportion of individuals who are isolated; and T is a binary treatment indicator. If an individual does not report, Pronyk et al. (2016) assume they remain infectious for 10 days.

The average time between symptom onset and reporting — $\bar{t}(T = 1)$ and $\bar{t}(T = 0)$ — can be calculated from data. In our sample, $\bar{t}(T = 0) = 4.73$ and $\bar{t}(T = 1) = 4.97$; we cannot reject the null that these are equal (see Table E.5).

Pronyk et al. (2016) assume a baseline reporting rate of 50 percent from mid-November to January, which is also the period that the disease was a major threat in our study area. 50 percent is consistent with other estimates, though it may understate the extent of under-reporting; the CDC's initial estimate was 40 percent.[A14] Going forward, assume $r(T = 0) = 0.5$ and $r(T = 1) = r(T = 0) \cdot \tau$.

Assuming the (initial) stock of Ebola cases is balanced across treatment and control, then $\tau = y(T = 1)/y(T = 0)$, where $y(T)$ is the number of reported cases and can be calculated from our data. Our estimates in Table 3 imply that $\hat{\tau} = (0.281 + 0.173)/(0.281) = 1.62$.

With these quantities in hand, we can calculate $D(T)$:

$$D(T = 0) = (4.72)(0.5) + 10[1 - (0.5)] = 7.46$$
$$D(T = 1) = (4.97)(0.5 \cdot 1.62) + 10[1 - (0.5 \cdot 1.62)] = 5.93$$

This implies that treatment generated a 19 percent reduction in R_0. Pronyk et al. (2016) estimate that CCCs contributed to a 13–32 percent reduction in R_0.[A15]

[A13] https://ajph.aphapublications.org/doi/suppl/10.2105/AJPH.2015.303020/suppl_file/web+appendix+r2.docx

[A14] https://stacks.cdc.gov/view/cdc/24901

[A15] Their estimate is likely conservative, as they do not incorporate how Community Care Centers affect reporting rates.