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

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

October 25, 2020

**Abstract**

Skepticism about the quality of health systems and their consequent underuse are thought to contribute to high rates of mortality in the developing world. The perceived quality of health services may be especially critical during epidemics, when people choose whether to cooperate with response efforts and frontline health workers. Can improving the perceived quality of health care promote community health and ultimately help to contain epidemics? We leverage a field experiment in Sierra Leone to answer this question in the context of the 2014 West Africa Ebola crisis. Two years before the outbreak, we randomly assigned two interventions to government-run health clinics—one focused on community monitoring, and the other conferred non-financial awards to clinic staff. Prior to the Ebola crisis, both interventions increased clinic utilization and patient satisfaction. Community monitoring additionally improved child health, leading to 38 percent fewer deaths of children under five. Later, during the crisis, the interventions also increased reporting of Ebola cases by 62 percent, and community monitoring significantly reduced Ebola-related deaths. Evidence on mechanisms suggests that both interventions improved the perceived quality of health care, encouraging patients to report Ebola symptoms and receive medical care. Improvements in health outcomes under community monitoring suggest that these changes partly reflect a rise in the underlying quality of administered care. Overall, our results indicate that promoting accountability not only has the power to improve health systems during normal times, but can also make them more resilient to emergent crises.

**Keywords:** disease control & prevention, Ebola, epidemic containment, government accountability, health systems, monitoring, non-monetary incentives, public service delivery, community monitoring

**JEL Classifications: I18, J33, M52, O15**

---

*This study uses a field experiment implemented in collaboration with the Government of Sierra Leone's Decentralization Secretariat and Ministry of Health and Sanitation, the World Bank, the International Rescue Committee, Concern Worldwide, and Plan International. We thank the Njala University Museum and Archive for sharing the de-identified data on Ebola records. We also thank Innovations for Poverty Action for collecting the original survey data, and the respondents and the accountability team for donating their time. Gieltje Adriaans, Ali Ahmed, Carolina Bernal, Alix Bonargent, Fatu Conteh, Afke de Jager, Sarah Dykstra, Caroline Fry, Kevin Grieco, Anne Karing, Magdalena Larreboure, Anthony Mansaray, Josh McCann, Niccolo Meriggi, Nick Otis, Moritz Poll, Jimena Romero, Mirella Schrijvers, and Samantha Zaldivar Chimal provided excellent research assistance. For comments, we thank Rachel Glennerster, Sendhil Mullainathan, Dan Posner, Manisha Shah, Macartan Humphreys and workshop participants at Berkeley, Columbia, LSE, UC San Diego, Zurich, Yale, Northwestern, Norwich, Amsterdam, Rotterdam, WZB Berlin, Wageningen, EGAP Nairobi, FHI360, UCLA, the World Bank's ABCA, and APSA. We gratefully acknowledge funding from USAID-DIV, the International Growth Centre, AFOSR grant #FA9550-09-1-0314, NWO grant #451-14-001, ESRC grant #ES/J017620/1, the Royal Netherlands Embassy in Ghana, and UCLA's California Center for Population Research. All errors are our own. Analysis of the survey-based outcomes was registered on the AEA registry: https://www.socialscienceregistry.org/trials/2085

†UCLA, Luskin School of Public Affairs (darinc@luskin.ucla.edu)

‡University of Chicago, Harris School of Public Policy and NBER (odube@uchicago.edu)

§Stockholm University, Max Planck Institute for Collective Goods, and Institute for Industrial Economics (johannes.haushofer@ne.su.se)

¶University of California, Berkeley, Center for Effective Global Action (bilal.siddiqi@berkeley.edu)

‖Wageningen University and Research (maarten.voors@wur.nl)

# 1. Introduction

Over 8 million people die annually in low- and middle-income countries from treatable conditions, generating human suffering and $6 trillion in economic losses (Kruk et al. 2018). These deaths are especially tragic because treatment is often not only possible, but also cheap and accessible (Deaton 2013). Yet, potentially life-saving health services remain underutilized due, in part, to the low perceived quality of health care (Dupas 2011; Banerjee, Deaton and Duflo 2004; Das et al. 2016). In a 2018 survey across 12 countries, more than half of the patients surveyed report that they did not seek necessary medical care in the previous year because they doubted the quality of their health system (Kruk et al. 2018). This not only frustrates the treatment of endemic diseases, but may also undermine the containment of emergent epidemics. Curbing epidemics requires compliance with public health directives related to, for example, testing and quarantine. As evidenced by the outbreaks of COVID-19, Zika, and Ebola, epidemics and pandemics recur with devastating local and global effects.

How can the quality of health care be improved? And do programs that achieve this goal under normal conditions also work when crises hit? We address these questions in the context of Sierra Leone, a country whose chronic health problems were compounded by the West Africa Ebola crisis. In September 2014, the World Health Organization (WHO) described the epidemic as "the most severe acute public health emergency seen in modern times." (WHO 2014a). 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, with Sierra Leone accounting for roughly half of those cases and just under 4,000 deaths (CDC 2019).

Prior to Sierra Leone's Ebola outbreak, we designed a large-scale field experiment to evaluate two programs intended to improve the utilization of government-run clinics and the quality of care delivered at these facilities. The timing of our study enables us to examine the programs' effects both under "normal conditions", and during the ensuing Ebola crisis. Endline surveying concluded in June 2013; the first Ebola case was reported in May 2014 (see Figure 1). We can, thus, observe whether the interventions contribute to the health system's resilience—the capacity to respond to crises and changing population needs that we observe only when a system faces an adverse shock.

We randomly assigned 254 clinics to one of the two interventions or control, in partnership with the Government of Sierra Leone (GoSL) and three international NGOs.[1] The first intervention, community monitoring (CM), provided patients with information and a public forum to monitor frontline health workers. Modeled on a program evaluated by Björkman and Svensson (2009), the intervention distributed scorecards

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

to rate local health services, and convened interface meetings between community members and health workers to discuss these ratings and develop "joint action plans" to improve service delivery. The second intervention provided non-financial awards (NFA) to both the best and most improved clinic in each district. Clinic staff were encouraged to develop action plans, and the winning clinics received wall plaques and letters of commendation from the district government. Neither program provided resources to clinics; rather, they intended to motivate health workers to supply higher quality care under existing resource constraints. The programs draw upon insights from personnel economics about how to motivate difficult-to-monitor frontline workers (Finan, Olken and Pande 2017). One strand of this literature focuses on non-monetary approaches, recognizing that performance pay may not be financially feasible or could crowd-out intrinsic motivation (Besley and Ghatak 2005; Bénabou and Tirole 2003; Dixit et al. 2002). Organizations can improve workers' performance by harnessing social incentives that arise from interactions between providers and clients, or among providers themselves (Ashraf and Bandiera 2018): the CM program empowers citizens to monitor providers and sanction those that under-perform (Mansuri and Rao 2003), while the NFA program engenders competition among health workers to improve service delivery (Besley and Ghatak 2005).

Prior to the Ebola crisis, we find that both interventions improve the perceived quality of health care. We define perceived quality of care as encompassing both the actual quality of care, as well as beliefs about the care provided at clinics. We cannot always disentangle changes in objective and perceived quality; we report evidence consistent with changes in both. We find, for example, that both CM and NFA increase the general utilization of health clinics. CM additionally improves maternal utilization—the probability of delivering a child in a health facility increases by 11 percent. We regard utilization as a revealed preference measure; our results, thus, suggest that individuals act on perceived improvements in the quality of care. In both treatment arms, we also find greater patient satisfaction, including satisfaction with the performance of health workers. These results are again consistent with improvements in the perceived quality of care provided by staff at program clinics.

We do not observe patient-provider interactions and, thus, cannot directly measure the quality of administered care. We do, however, measure health outcomes, which we expect correlate positively with quality. The CM program produces substantial improvements in child health outcomes: the likelihood of under-5 death in the household declines by 38 percent. These effects are similar in magnitude to Björkman and Svensson (2009), who find a 33 percent reduction in under-5 mortality in Uganda. These improvements could reflect increased utilization as individuals seek treatment and care. However, both CM and NFA increase utilization, while CM alone bolsters child health, which suggests additional improvements in the quality of administered care, particularly under community monitoring.

We then assess the effects of these programs during the ensuing Ebola epidemic. Ebola containment efforts emphasize early isolation and treatment. Yet, fears about sub-standard care and a lack of confidence in health workers deterred symptomatic patients in Sierra Leone from visiting clinics. Instead, individuals

hid sick family members and evaded testing and contact tracing efforts (Abramowitz et al. 2016, 24).

To test whether our interventions contribute to the health system's resilience, we ask whether they affect reporting of Ebola cases. We use a de-identified database maintained by the GoSL and CDC to construct weekly counts of tested and confirmed patients in small administrative units called sections. We focus on the 160 sections that contain a single clinic from the experimental sample, which permits unambiguous coding of each section's treatment status. Pooling the two interventions, we estimate that they substantially increased reporting, by just over 60 percent. While we cannot reject the null hypothesis that the interventions have statistically equivalent effects, qualitatively, we see a larger increase in reported cases in sections containing clinics under CM.

We attribute increased case counts to reporting behavior, not Ebola transmission. The programs increase all types of cases—confirmed cases, and cases in which patients test negative for the virus. We also rule out nosocomial transmission (i.e., exposure to infected patients in a clinical setting) in 99 percent of cases, based on the timing of symptom onset and reporting. We thus interpret the increase in reported cases (including confirmed cases) as a critical step toward containment: a back-of-the-envelope calculation (per Pronyk et al. 2016) suggests that this increased reporting reduces the virus's reproduction rate ($R_0$) by around 19 percent.

These findings align with our results prior to the crisis: improvements in the perceived quality of care encourage reporting during the epidemic. In particular, we show that general utilization, satisfaction with public health workers, and confidence in the effectiveness of Western ("white-man") medicine relative to traditional healers all significantly increase in treatment areas in the 160 clinics used in our Ebola analysis. We combine these measures into a perceived quality of care index. Instrumenting that index with our randomized treatment assignment (per Kling, Liebman and Katz 2007), we find that a one standard deviation change in the perceived quality of care increases Ebola reporting by 0.43 cases per section-week. Separating the two interventions, we find that CM has larger (first-stage) effects on our quality of care index, consistent with its larger effects on reporting noted above.

We find no evidence of enhanced disease surveillance in areas with program clinics, further supporting the view that the primary effect of the programs is on reporting behavior. Sections with program clinics do not host more facilities specializing in Ebola care, and there are no differences in laboratory testing or case workers. The treatments also do not increase contact tracing efforts (the process of identifying recent contacts to flag at-risk individuals); in fact, there is more contact tracing in sections with control clinics. We also find no evidence that geographic spillovers—the movement of patients from control to treatment sections—amplify our effects.

Beyond reporting, we examine mortality among Ebola patients. In sections with CM clinics, we ob-

serve a decline in mortality: one patient dies for every ten that report, compared to one in four in sections with control clinics. This result is conditional on reported cases. Thus our estimates, again, suggest that CM generates benefits through a channel beyond utilization (e.g., through changes in the quality of administered care). Because improvements in health outcomes are concentrated in CM clinics, both under normal and crisis conditions, the direct community involvement under CM may spur a larger and sustained change in providers' behavior and the resulting quality of health services.

Our results highlight that non-monetary approaches can improve the perceived quality of health care; and that these improvements strengthen health systems, bolstering their resilience to crises. These points contribute to related literatures on how to improve service delivery and build trust in public services.

A large literature addresses the challenges of motivating frontline bureaucrats responsible for delivering services. Community monitoring has been employed across a variety of 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. In the health sector, community monitoring appears to have larger effects in contexts with poor baseline health outcomes and services, such as in Uganda in 2005 (Björkman and Svensson 2009), India (Mohanan et al. 2020), and our study in Sierra Leone. Encouragingly, Björkman Nyqvist, de Walque and Svensson (2017) find that these effects persist: following up on Björkman and Svensson (2009), they find lasting treatment effects on health outcomes over the longer run. By contrast, community monitoring may not work as well when baseline health conditions are better (for a recent study in Uganda, see Raffler, Posner and Parkerson 2019).

Our results also confirm prior findings that show non-financial awards can boost performance among mission-oriented workers (Ashraf, Bandiera and Jack 2014) and in other settings (Kosfeld and Neckermann 2011; Ball et al. 2001; Markham, Scott and McKee 2002). Gains may be smaller if providers learn the formula for allocating non-financial awards and distort their effort toward rewarded tasks (Glewwe, Ilias and Kremer 2010). To avoid this issue, we did not disclose the metrics used to rank clinics in our NFA intervention. Community monitoring and non-financial awards represent only two approaches to harnessing social incentives and improving health care in developing countries; Dupas (2011) and Dupas and Miguel (2017) provide reviews.[2]

On the demand-side, a growing body of work finds that trust affects the utilization of public services, particularly health care. Fear and distrust deter patients from utilizing health systems over a long horizon (Alsan and Wanamaker 2017; Lowes and Montero 2018). Patients' trust may be particularly important amid public health crises, when they face choices about whether to voluntarily report for medical testing or honor

[2]Other studies examine the effects of community health workers (Björkman Nyqvist et al. 2019); financial incentives (Miller et al. 2012; Olken, Onishi and Wong 2014; Singh and Mitra 2017); career opportunities (Ashraf et al. 2020); technological monitoring combined with financial incentives (Banerjee, Duflo and Glennerster 2008); and social signaling among patients (Karing 2019).

a quarantine. Our findings reinforce work in Liberia (Blair, Morse and Tsai 2017; Morse et al. 2016; Tsai, Morse and Blair 2019) and the Democratic Republic of Congo (Vinck et al. 2019), which finds that trust in government affected clinic utilization during those countries' Ebola crises. This research is also echoed in commentary on the COVID-19 pandemic, with experts arguing that distrust of public health officials undermines containment efforts.[3]

The rest of our paper is structured as follows. Section 2 describes the study context, 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 findings under normal conditions and the longer run effects under the Ebola crisis, and additionally discusses cost-effectiveness. The final section concludes. The Appendix figures and tables that we reference are included in an Online Appendix.

# 2. Health Care in Sierra Leone

## 2.1 Background

In 2010, Sierra Leone had the highest maternal mortality rate in the world, at 13.6 deaths per 1,000 live births, and under-5 mortality stood at 162.8 deaths per 1,000 live births. Appendix Figure A.1 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. Western-style health care is provided primarily through government-run clinics and hospitals; private and NGO-sponsored facilities are scant (Denney and Mallett 2014). Government facilities operate alongside traditional village birth attendants and healers. Our study focuses on primary health clinics—the first points of contact for patients in towns and villages—that each serve populations of 500 to 10,000 (UNICEF 2014). These clinics typically focus on maternal and child care, providing services such as antenatal care, supervised deliveries, postnatal care, family planning, growth monitoring for under-five children, and immunization. In addition, there is some focus on health education and management of minor ailments, as well as referral of more serious medical conditions to larger facilities (MOH 2017).

In an effort to reduce child and maternal mortality, the GoSL launched a free health care initiative in 2010, removing fees for pregnant and lactating women and children under the age of five. The policy simultaneously increased pay for government health care workers; at the time, 30–50 percent of staff did not

[3] 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."

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.

Primary health clinics continued to operate during the Ebola crisis: a UNICEF (2014) facility survey in October 2014 (four months after the first confirmed case in Sierra Leone) found that only 4 percent of clinics were closed. In addition, clinics remained largely accessible: the GoSL implemented short lockdowns, most prominently a three-day nationwide quarantine between 19 and 21 September 2014 that banned all travel. We know of no additional travel bans within our study area that impacted clinic access. Levy et al. (2015) report that "early assessments [from October 2014] found that many [Ebola] patients were first seeking care at local [clinics]." Concerned that these clinics lacked the training and equipment to properly isolate and care for Ebola patients, clinic 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 Appendix Table E.1); by late December 2014, training had reached 98 percent (Levy et al. 2015). Case studies suggest that clinic staff and community health workers were providing "no-touch" treatment for dehydration and fever, and engaged in social mobilization and disease surveillance (Vandi et al. 2017). While training and the disbursement of protective equipment filled important knowledge and resource gaps, UNICEF's (2014) survey found that 90 percent of clinic staff felt that fear and misconceptions were "the main challenge confronted by the health system in fighting Ebola."

## 2.2 Interventions

In addition to removing cost barriers and severe resource constraints, as part of the free health care initiative, the GoSL saw a need to strengthen incentives for front-line health care workers. Absent incentives tied to service delivery, the government worried that nurses would miss work or continue to charge illegal fees or inflated drug prices—barriers to service provision that the free health care initiative intended to eliminate.

With World Bank support, the GoSL contracted with three international NGOs to implement two interventions in 170 clinics across four districts. Plan International worked in Bombali district, Concern Worldwide in Tonkolili district, and the International Rescue Committee worked in Bo and Kenema districts.[4] The four districts bisect Sierra Leone from North to South (see Appendix Figure A.2) and cover just over 30 percent of Sierra Leone's population.

[4]Implementation by multiple international NGOs with broad development portfolios suggests that the interventions did not require a local implementer or one specialized in health care.

### 2.2.1 Community Monitoring (CM)

The community monitoring intervention was modelled on Björkman and Svensson's (2009) "Power to the People" approach implemented in Uganda in 2005. The intervention attempts to mobilize "client power," providing patients with information and a forum to demand accountability from frontline staff (The World Bank 2003). It 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 followed a four-step protocol. First, trained facilitators organized meetings with clinic staff and shared scorecards rating local health problems. The scorecard included 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 were constructed from administrative data provided by the Ministry of Health and Sanitation and compared to the district average so as to prompt discussion. Clinic staff were then invited to share their concerns and frustrations with the community. For example, nurses frequently complained that community members did not visit the clinic when they were sick, mothers opted against in-patient deliveries, and parents failed to complete the vaccination courses for children.

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

Third, interface meetings brought together community members and clinic staff. Facilitators guided a discussion in which both sides had the opportunity to articulate the complaints and concerns raised in the earlier meetings. The facilitators then assisted clinic staff and community members to formulate a joint action plan that specified the actions each party would take to improve health care. Facilitators worked with both sides to specify a time-frame and assign a responsible "point person" for each component of the action plan. Meetings concluded with community and clinic representatives signing the plan. Several of the most common problems cited in the plans relate to utilization, and listed a range of actions that users and providers jointly agreed on to target this outcome. For example, health facility staff were 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 agreed to seek care at the clinic more promptly and consistently for their health needs. After the meeting, facilitators left a copy of the action plan with the clinic and representatives from each village.

Finally, facilitators held follow-up meetings three, five, and nine months after the initial interface meet-

ing to revisit the action plan and monitor progress. These meetings were held jointly with the community and clinic staff, and each side rated the extent to which the other side had made progress on their commitments. The research team monitored almost all CM clinics at some stage of the intervention.[5]

### 2.2.2 Non-financial Awards (NFA)

The non-financial awards intervention set up district-wide competitions among clinics. Clinics were ranked at baseline and endline, using data collected at clinics. Awards were given to both the best and most-improved clinics within each district. The second award helped encourage staff to improve performance at clinics with low baseline rankings, who might have otherwise been demotivated. In total eight awards were allocated across the four study districts; and just under 10 percent of the 85 NFA clinics received an award.

The average clinic has just over two staff members, and this small size ameliorates free-riding problems that might otherwise arise in a competition that awards clinic-wide outcomes, rather than individual effort. Key performance indicators included measures of utilization for antenatal care, childbirth, and vaccinations, as well as users' experiences, including absenteeism, staff attitude, and charging fees for free services. Importantly, these indicators were not revealed publicly to avoid having staff reallocate their effort toward these tasks at the expense of other important tasks.

To encourage truthful reporting of indicators, clinics across all treatment groups were informed that their patient registers would be audited at baseline and endline, and clinics with fraudulent entries would be disqualified from the competitions. Each audit involved randomly selecting 30 patients from the clinic register (corresponding to 15 patients per study community) and visiting each individual to verify their recorded visit date and purpose. None of the audits uncovered ghost patients or manipulated entries in the clinic register.[6]

The implementing NGOs took a number of steps to build broad awareness of the competition. First, they met with district health officials to explain the competition. Second, they advertised the competition extensively in and around the clinics: posters were placed inside the clinic and at high-traffic locations,

[5] We randomly selected and observed half of the interface meetings. Meetings typically lasted 3–4 hours; average meeting attendance ranged from 52 people in Kenema district to 68 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). We also monitored the three-month follow-up meeting for nearly all clinics where we did not observe the initial interface meeting. In the three-month follow-up, average meeting attendance was only slightly lower than at the first interface meetings.

[6] 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 the endline, we sampled clinic registers from clinics in all study clinics; however, to reduce data collection costs, we only visited patients to verify details for NFA clinics to ensure that awards were handed out correctly. Verification in CM and control clinics would not differentially affect reporting, since the endline data was collected prior to the second round of verification, at a time when all clinics still expected to be audited.

such as schools and chiefs' homes. Third, the NGOs met with each of the clinics individually to explain the competition and answer relevant questions. During this meeting, they encouraged clinic staff to develop action plans to identify opportunities to improve service provision, without divulging which indicators would be used for the actual rankings. Finally, they held follow-up meetings at the clinics three and six months after the initial meeting to remind staff about the competition. The research team monitored at least one meeting in 49 out of 85 NFA clinics.

Winners were not announced until after the endline survey. Winning clinics received a wall plaque to display inside the clinic at a public ceremony, and all staff at winning clinics received letters of commendation from district health officials.

The awards were "non-financial" from the government's perspective, as they did not involve any monetary compensation. Workers could, nonetheless, have associated winning with a longer-term financial payoff. For example, they could have anticipated that being on staff at a winning clinic would lead to promotions or transfers to attractive locations. We are agnostic about which element of the award, recognition or career concerns, motivated workers.

# 3. Design and Methods

## 3.1 Sampling

### 3.1.1 Clinics, Communities, and Households in Full Experimental Sample

The districts in our study include 318 primary health clinics. We sampled 254, 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 in our sample is 10 kilometers. In Sierra Leone, each clinic has a defined catchment area (a roughly 3-kilometer buffer around a clinic) that prioritizes the communities it serves. Individuals are also administratively assigned to a specific clinic, which, combined with high travel costs, discourages the use of more distant clinics. At baseline, the average clinic in our sample had just over two staff members present, reported being open six days per week, and saw roughly 450 patients per month. 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 randomly sampled two communities from each clinic's catchment area. As shown in the Consort

Diagram in Figure 1, 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 community and administered a shorter user-feedback survey focused on recent health episodes, service provision, and satisfaction. For the 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. 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 lived in homes with mud floors and wooden walls, 24 percent had no toilet facility, another 58 percent used a pit latrine, only 20 percent owned a mobile phone, and 62 percent had no formal education.

#### 3.1.2 Blocking and Randomization

We grouped the 254 clinics in our sample into matched triplets using Greevy and Beck's (2016) non-bipartite matching algorithm. Clinics in a triplet fall within the same district and exhibit similar levels of utilization and performance at baseline.[7] Blocking on matched triplet, we randomized 84 clinics into control, 85 into CM, and 85 into NFA.

[Figure 1 about here.]

#### 3.1.3 Sections in Ebola Sample

We are not able to associate Ebola cases with specific clinics or geolocate them accurately to clinic catchments. The smallest unit to which we can confidently geolocate cases is the section—the smallest administrative unit in Sierra Leone, which is typically just under 40 sq km in size and has fewer than 2,500 residents according to the 2004 census (see Appendix Figure A.2(c) for a map of section boundaries). Therefore, we aggregate Ebola cases to the section level. We discuss this procedure further in Section 3.2.2 and provide greater detail in Appendix Section E.2.

The 254 clinics in our experimental sample fall into 205 sections. Of these 205 sections, 45 include multiple sample clinics. In our primary Ebola analysis, we restrict attention to the remaining 160 sections that contain a single study clinic and, thus, a unique treatment assignment. Figure 2(a) maps the 205 sections, with those included in the primary Ebola sub-sample shown in darker gray. Within the 160 sections in our

[7] We exactly match clinics by district and clinic type (maternal and child health post or community health post). We then select matches based on the Mahalanobis distance between eight indicators specified by the Ministry of Health and Sanitation: 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.

primary Ebola sample, 54 are control, 46 CM, and 60 NFA. As a robustness check, we also analyze the Ebola data using a dose-response model, which uses all 205 sections and measures dosage as the proportion of study clinics in each section that receive either treatment.

[Figure 2 about here.]

Since we only sampled 254 out of 318 total primary health clinics, even sections with one sample clinic can contain a non-sample clinic and, thus, more than one total clinic. However, additional non-sample clinics are rare: among the 160 sections in the primary Ebola sample, on average, the share of sample clinics out of total clinics is 94 percent. Moreover, this is balanced across treatment and control sections (see Appendix Table E.13). This suggests that Ebola cases can largely be attributed to the experimental clinic.

## 3.2 Data Collection

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

Baseline surveys were administered in September 2011, and endline surveys in May and June of 2013 (see Figure 1 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 in each community regarding amenities, relations with the clinic, and community development; and third, household surveys which captured attitudes, behaviors, and outcomes related to health and economic well-being.

We filed an analysis plan to examine the survey outcomes at the AEA RCT registry.[8] The plan defines ten outcome families, including sub-components that comprise each family. We flag and explain any subsequent deviations in Appendix Section B.1.

Each outcome family represents a set of variables aggregated using control group-standardized indices per Kling, Liebman and Katz (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 $\widetilde{y}_i = \frac{1}{K}\sum^{K}\left(\frac{y_{ik}-\mu_{0k}}{\sigma_{0k}}\right)$,

[8] AEARCTR-0002085: https://www.socialscienceregistry.org/trials/2085, March 2017. This plan was filed after data collection and preliminary data analysis conducted for a brief report to the GoSL, which was a contractually required deliverable of the project. For the report we analyzed outcomes agreed upon at the beginning of the study: institutional delivery, antenatal care visits, immunization, illegal fees, nurse absenteeism, staff attitude, maternal and under-five mortality, utilization, and anthropometric outcomes. We did not examine other outcomes from the household data, or any outcomes from the clinic or community data.

where $\mu_{0k}$ and $\sigma_{0k}$ are the estimated control-group mean and standard deviation for outcome $k$ in family $K$. Our estimates for these families thus represent standard deviation changes relative to the control group. Following Kling, Liebman and Katz (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. For that reason, 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 and Liebman (2004) and Casey, Glennerster and Miguel (2012) and aggregate treatment effects across the sub-components of each family using seemingly-unrelated regressions (SUR). These results (reported in Appendix Section D) are qualitatively similar across all specifications.

Below we describe each outcome family; Appendix Section B.1 provides additional detail on each family's sub-components, and Appendix Table B.1 includes descriptive statistics of each variable at endline.

1. **General utilization** measures the number of episodes in which individuals seek care at a Western-style clinic, including in response to four of the most common health needs addressed at primary health units—childbirth in the past year, antenatal or postnatal care, vaccination, or any illness or injury, as well as a residual category of any other type of consultation in the past month.[9] While most utilization occurs in response to specific health needs (as regular health check-ups are not common in our setting), the residual category helps generate a comprehensive measure of utilization. Utilization of a Western-style clinic reflects the decision to seek care at a formal clinic, rather than visiting a traditional healer or spiritual leader or forgoing any type of care.[10]The Western-style clinics utilized by respondents are overwhelmingly government-run clinics; utilization of private or NGO-run clinics constitutes a small share of utilization (3 percent).

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 clinic.[11]

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

[9]We use the number of visits to Western-style clinics instead of the proportion of visits out of reported health episodes, as the count captures both changes in the propensity to report a health episode and the propensity to seek care at a clinic conditional on having reported an episode (see Appendix Section B.1 footnote A1).

[10]Our analysis plan specified examining utilization of traditional religious healers. However, due to an error in survey design we do not have complete utilization data for traditional healers for all health episode types (see footnote A2 in Appendix Section B.1). We therefore focus on utilization of Western-style clinics in the main results. We do however have utilization data for both Western clinics and traditional healers for one specific episode type, namely illness/injury episodes. We conduct robustness checks using this episode type alone in Appendix Tables D.21 and D.22.

[11]Some outcomes within families (e.g., ANC/PNC visits) are themselves indices; for these we continue to use the control-group standardized indices described above.

ratio.[12] 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 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 of the sample)—satisfaction with the care they received.[13] Among 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 all types of health episodes, so we average responses across individuals in a household when multiple episodes are reported.

5. **Clinic organization and services** 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.

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.

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

[13] 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 an index measured at the household level that includes 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.

### 3.2.2 Ebola Case Data

We rely on a de-identified version of the Epi Info Viral Hemorrhagic Fever (VHF) database, which was the primary data management system used to track the Ebola outbreak in Sierra Leone.[14] 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) describe it as "the most comprehensive epidemiologic and laboratory data on Ebola cases available in Sierra Leone." The VHF compiles patient information, their lab results, and whether they died. Patients could enter the VHF through walk-in visits to health centers, as well as surveillance activities (e.g., contact tracing) (Owada et al. 2016). As noted above, the VHF reflects reported cases, rather than actual Ebola incidence—a particularly important outcome for stopping contagion and containing the epidemic (Enserink 2014).

We use information on patients' residences to geocode cases to sections. (The location of symptom onset is recorded for only a subset of cases; when it is not missing, it matches the patient's residence for over 90 percent of cases.) We aggregate cases to sections rather than villages, towns or smaller geographic units owing to several features of our geocoding procedure (see Appendix Section E.2). First, many villages in Sierra Leone do not have recorded names; when patients report their community of residence, they tend to name better-known towns, rather than their village. Often this will be the name of a central or headquarter town of the section, which are formal administrative units. By aggregating cases to the larger administrative unit, we avoid measurement error that arises from attributing cases 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 to which we can match patients' residences: 85 percent of census enumeration areas (which are just 7 square kilometers on average) contain one or zero placenames. By contrast, the average section (averaging 40 square kilomteres) contains eight geolocated placenames; and 94 percent of sections contain more than one placename. Our geocoding protocol does not introduce imbalance: we find that treated and control sections do not differ significantly in terms of having more or longer placenames (see Appendix Table E.2).

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

In the 160 sections that constitute our primary Ebola sample, the VHF includes 2,045 case entries, which are classified into four types: 1,623 **negative** cases where Ebola has been ruled out; 269 **confirmed** cases; and two residual categories that are never confirmed with lab tests: 134 **suspected** cases which display Ebola symptoms and/or have had contact with potentially infected individuals or animals and 19 **probable** cases which 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.[15] Given our interest in reporting, our main dependent variable is the count of **total cases** (the sum across the four case types) aggregated to the section-week. We use the date when a case is first entered in the VHF database to determine the week. Appendix Table E.3 presents descriptive statistics for total and confirmed cases.

## 3.3 Manipulation Checks and Balance

### 3.3.1 Manipulation checks

We take several steps to verify that implementation matched our randomized assignment. First, a monitoring team visited 95 percent of the CM clinics and 58 percent of the NFA clinics to verify that activities matched treatment assignment and conformed to protocols. At every site, activities matched the assigned treatment. Second, one of the implementing NGOs provided detailed data on all activities (including dates), enabling us to cross-check compliance throughout Tonkolili District. We uncover no deviations. In addition, the implementing NGOs implementation budgets were tied to the number of treated clinics in their districts; as a result they had neither the incentive or resources to target additional (control) clinics.

Finally, as specified in our analysis plan, we asked survey respondents whether key program activities took place in their communities or clinics. In Appendix Table C.1, we find that 86 percent of leaders in CM communities report "meetings held by IRC, Plan or Concern to discuss how the clinic and community can work together to improve service delivery in this community." This is roughly double the rate compared to control communities and suggests broad awareness of interface meetings in CM communities. Note, however, the high control mean (over 40% of control communities also report a community meeting). This likely reflects confusion, as NGOs commonly convene community meetings across rural Sierra Leone. NFA communities also report an increase on this measure (of 12 percentage points), albeit significantly less than CM. Given the monitoring we describe above, we do not attribute this to contamination. Rather, the NFA protocol also involved meetings convened by these NGOs to develop action plans to improve

[15] 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 overlooked or not entered into the VHF.

service delivery in the community. The leaders answering the community survey in NFA communities may have (understandably) responded affirmatively; some of these leaders actually participated in the meetings convened at NFA clinics.

Appendix Table C.2 offers a similar story: over 81 percent of staff at NFA clinics report participating in a competition, which is five times the rate among control clinics. Yet, smaller shares of staff at clinics in both control and, to a greater extent, CM also report competing. (The difference across treatment arms is 46 percentage points, which is substantial and statistically significant.) We visited nearly all CM clinics and found no NFA programming. Thus we attribute these responses to mis-interpretation: staff in CM clinics may have interpreted the scorecards (and the comparison to other clinics in the same district) as inviting competition across facilities.

### 3.3.2 Balance in Full Experimental Sample

Appendix Table C.3 reports balance across pre-specified covariates. Most variables are individually balanced across treatment arms. We find that the number of injuries or illnesses reported is lower in both CM and NFA relative to control; and in CM, household size is slightly smaller, there is lower trust of Village Health Committees (VHCs) and fewer households report a recent childbirth. However, 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 cellphone coverage, and individuals are less likely to belong to the Temne ethnic group, less likely to believe what a doctor told them, and have a higher level of educational attainment.[16] Given two treatment arms and a control group, we use a multinomial logit model to assess whether the baseline covariates jointly predict treatment assignment. At the bottom of the balance table, we report $p$-values from Chi-squared tests of joint orthogonality, following Özler et al. (2018). These tests suggest that the covariates together are not jointly significant. We also report results where we control for baseline imbalance in Appendix Table D.24.

### 3.3.3 Balance in Ebola Sample

Appendix Table E.4 reports balance checks for the 160 sections in our Ebola sample. Some of the imbalance observed in the full sample carries over to into this subset, while a small number of variables are imbalanced in the subset but not in the main sample (i.e., CM and NFA communities are less likely to have a prominent

[16] We observe imbalance when we analyze the pre-specified variable "Is there phone coverage within 1 mile from the community" from the community survey. However, we do not see imbalance when we analyze a closely related question: "Is there phone coverage in a one-mile radius around the community where the facility is located." Nor do we observe imbalance on two questions of mobile phone ownership from the clinic and community surveys. These results (available upon request) suggest that there are no systematic differences in access to communications in treatment versus control clinics.

village member in the household, while CM communities are more likely to have a mortorable road, lower trust and more members of the Temne vs. Mende ethnic group.) But again, Chi-squared tests of joint orthogonality presented at the bottom of this table indicate that the covariates are not jointly significant in the Ebola sub-sample. We also report results where we control for baseline imbalance when examining Ebola outcomes in Appendix Table E.21.

## 3.4 Specifications

### 3.4.1 Survey Outcomes

Our main specification is the following ANCOVA-type model:

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

where $y_{ivc,EL}$ is the outcome of household (or individual) $i$ in village $v$ in clinic catchment $c$ at endline ($EL$). $\alpha_b$ 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. If this variable is missing for a given village, we incorporate imputed values and a separate indicator variable for these observations, which controls directly for the imputation effect while enabling us to retain these observations in our sample. Thus we estimate Missing Indicator ANCOVA models. We use the village average because due to cost considerations, our baseline survey included a smaller sample of households. Moreover, some outcomes are only defined for individuals who recently experienced a given health episode. Some households surveyed at endline that experienced relatively infrequent health episodes, such as childbirth, would likely not have also experienced the same episode at baseline. For both reasons, controlling for a household's baseline outcome would reduce the size and representativeness of our sample, and we therefore use the village-level average. When $y$ is a sub-component of an outcome family, we use the family-level outcome to compute the baseline average.[17] We cluster our standard errors on clinic, the unit of randomization. We also estimate a variant of Equation 1 in which we combine the CM and NFA treatments into one pooled treatment indicator.

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

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

[17] 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 any of our conclusions.

These models include a single observation per clinic, removing the need to cluster standard errors on clinic as in Equation 1.

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 (FDR) within treatment arm (1) across outcome families and (2) across sub-components within each family.[18]

### 3.4.2 Ebola Outcomes

We assess the impact of the CM and NFA interventions on reported cases for the 160 sections in the Ebola sample (described in Section 3.1.3). We observe counts of reported cases in each section in every week from 10 August 2014 to 18 October 2015. We restrict attention to the period from September 2014 through April 2015, when Ebola transmission was a real threat in our study area; only three confirmed cases were reported during May and October 2015.[19]

Using this data, we estimate:

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

where $\alpha_b$ again represents the matched-triplet fixed effects; $\delta_t$ 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.[20]

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, number and population of bordering control sections, as well as number of proximate control sections with the same plurality ethnic group).

[18] In the analysis plan, we specified controlling for the FDR only across some families (denoted 'primary families') and, within those families, only across some sub-components. However, since we examine all outcomes, we take a more conservative approach and instead correct for multiple comparisons across all outcome families and, within each family, across all sub-components.

[19] In Appendix Table E.5 we extend the panel back to August 2014 and replicate our primary results from Table 3.

[20] 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 at the clinic level, and there is one clinic per section in the main Ebola sample, we do not cluster our standard errors in the cross-sectional models because section is both the unit of observation and treatment assignment.

# 4. Results

## 4.1 Effects Prior to the Ebola Crisis

Our tables follow a common format. Column 1 provides the control mean and standard deviation at endline, which by construction 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. Columns 3 and 4 separately estimate the average treatment effects for CM and NFA, respectively. Column 5 shows the difference between the average treatment effects 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. The tables in Appendix Section D aggregate across the sub-components of families using SUR.[21] These estimates, presented in the first row of every table, show qualitatively similar results to the mean effect indices. The remaining rows in these appendix tables also show treatment effects on individual sub-components.

Table 1 examines the interventions' effects on utilization, satisfaction, and health outcomes. Prior to the Ebola outbreak, both programs increase general utilization: the pooled treatment effect is 0.11 standard deviations, with statistically indistinguishable effects across the different two arms. Individuals in the control group used a Western-style clinic for roughly one (0.96) health episode (see Appendix Table D.1); the treatments increase utilization of such facilities by about 5 percent. When we focus attention on the utilization of government-run clinics in Appendix Table D.20, our effects increase to 7.2 and 5.9 percent for CM and NFA, verifying that the interventions boosted utilization of the targeted clinics.[22] We are only able to compare the utilization of Western-style clinics and traditional healers for one type of health episodes, illness and injuries. In Appendix Tables D.21 and D.22, we detect a shift (of roughly equal magnitude) away from traditional or religious care and toward Western-style clinics.

We interpret utilization as a revealed preference measure. Our results thus suggest that individuals responded to improvements in the perceived quality of care offered at CM and NFA clinics.[23] Perceived quality of care encompasses changes in actual quality, as well as beliefs about the care provided at clinics. Increased utilization (and the improvements in satisfaction we describe below) are consistent with improvements along one or both dimensions.

[21] Appendix Tables D.1–D.10 present treatment effects for each of the individual indicators. Appendix Tables D.11–D.19 repeat these analyses using the z-scored (i.e., control group-standardized) versions of the indicators.

[22] In robustness checks (available upon request), we find similar results if we measure utilization using the proportion of health episodes for which an individual utilized a Western-style clinic or a binary indicator for any utilization of a Western-style clinic.

[23] Spillovers do not provide a plausible explanation for increased utilization: our measure is based on household surveys, not clinic registers. If our respondents traveled to treated clinics for care, this would attenuate our estimates, as it would appear to increase utilization among households living near control clinics.

The CM arm also shows additional effects on maternal utilization: among women who gave birth in the year before the endline survey, maternal utilization increases by 0.18 standard deviation units. There is no equivalent effect for NFA. Appendix Table D.2 shows that the increase in maternal utilization is driven by more deliveries at Western-style clinics: the probability of giving birth in these facilities is 0.83 in control areas; CM boosts this rate by 9 percentage points (11 percent). We estimate no effect on antenatal and postnatal visits.

Consistent with the perceived quality of care improving, the third row of Table 1 shows increased patient satisfaction. We find similar effects in both arms. Pooling the treatments, satisfaction increases by about 0.10 standard deviations, largely driven by increases in respondents' satisfaction with their own health and the performance of health workers (see Appendix Table D.3).[24] We generally see high baseline levels of satisfaction, though 17.5 percent of respondents report they are somewhat or very unsatisfied with public health workers. Unsurprisingly, the programs' effects on this outcome reflect improvements among households with low baseline levels: splitting our sample into thirds using baseline responses, we find treatment effects on satisfaction with public health workers only in the bottom two terciles (results available upon request). All households are asked about their satisfaction with public health workers; thus these effects can arise from improved experiences at clinics, as well as hearing neighbors' positive assessments.

The quality of care administered at clinics (e.g., the time that nurses spend on diagnoses or treatment plans) may also improve. Unfortunately, patient-provider interactions are difficult to measure (for an exception, see Das et al. 2016). We look instead at child health outcomes, assuming that these respond to the actual quality of health care, not just parents' beliefs about the quality of clinics. The fourth row of Table 1 shows that CM leads to an improvement in health outcomes (0.17 standard deviations). This is driven by significant improvements in child health. As shown in Appendix Table D.4, the likelihood that a child under-5 dies in CM falls by 0.015 relative to the control mean of 0.03, a 38.4 percent effect. In addition, child weight-for-length increases by 0.16 z-score units and is significant at the 10% level, though this individual indicator loses significance after FDR adjustments. It is worth noting that the magnitudes of these effects are sizable and qualitatively similar to those uncovered by Björkman and Svensson's (2009) evaluation of community monitoring in rural Uganda. Finally, the effect size for vaccine completion is also large, corresponding to a 10 percent increase in vaccine completion, though the change is not statistically significant. Improvements in health outcomes could reflect increased utilization as individuals access more treatment and preventive care at clinics. However, general utilization increases in both arms, while health outcomes only improve under CM. In addition, the improvements in maternal utilization under CM reflect an increase in institutional deliveries over the last year rather than increases in ante- or post-natal visits (see Appendix Table D.2). Decisions by recent mothers (of which there are only 888) to deliver in clinics

[24] It is possible that the effects on satisfaction partly reflect social desirability bias in CM, where community members and clinic staff convened to discuss the state of local health care and health services. It is less clear why respondents in NFA would feel social pressure to report more satisfaction. The comparable effects across CM and NFA suggest that social desirability bias is unlikely to drive the estimated effects.

are unlikely to affect under-5 mortality among the much larger set of households included in our survey. These patterns suggest that an additional channel, i.e. greater effectiveness of health services, contributes to improved health outcomes under community monitoring.

[Table 1 about here.]

We next study effects on other families that could influence clinic utilization and health outcomes. In the top panel of Table 2, we examine the quantity of health services and community contributions to clinics, since either a larger menu of services or groundswell of support could draw in patients and improve their outcomes. The first row of shows no significant effects on health service delivery.[25] However, in Appendix Table D.5, we find divergent results for the indicators that measure the quantity of services versus those reflecting their quality. We find no effects on the availability of drugs, medicines in stock, or staff presence.[26] Yet, 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.[27] These effects, which point to more positive patient-provider interactions, suggest that the effectiveness of health services may also have improved with NFA, though the effects are not strong and do not suffice to improve health outcomes.

[Table 2 about here.]

When we examine effects on clinic organization and services in the second row of Table 2, we do not see any significant effects at the family level or for individual sub-components (see Appendix Table D.6). This is not surprising since neither intervention provided additional resources to clinics; the government's interest in the evaluation was understanding how to extract more effort from health workers under existing budget and logistical constraints. In addition, we find no significant change in community support: community members did not spend more time or resources on the clinic or its staff.

[25] 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.

[26] We observe a positive effect of NFA on absenteeism in Appendix Table D.5. 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. Appendix Table D.23 shows precise null effects on this outcome.

[27] There is a small (1.5 percent) and marginally significant increase in whether people would return to the clinic under CM. Note, that 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.

In the second panel of Table 2 we look at two other downstream outcomes that changed as a consequence of the interventions: the community's engagement with political and economic development efforts, and water and sanitation infrastructure. Both interventions led to improvements in community development and political engagement (CDPE): the pooled treatment effect is 0.23 standard deviations. This reflects two related changes: first, treated communities report more projects undertaken by local officials (e.g., chiefs), which were supported by voluntary labor; second, we see small ($< 1.5$ percent) increases in voter turnout in NFA (see Appendix Table D.8). Both findings are consistent with community members crediting local officials for efforts to improve public services. In CM, leaders played prominent roles at community meetings; in NFA, the program was advertised at chiefs' homes.

We also find improvements in water and sanitation in NFA communities in the final row of Table 2. These effects arise from households in NFA accessing better sources of drinking water: they report increased use of mechanical wells and boreholes, and decreased use of natural springs or water transported in jerrycans (results available upon request). These effects are consistent with different potential mechanisms—for example, discussions on strategies for improving health in the NFA clinics, and/or greater mobilization around development projects, may have led to higher prioritization of securing access to safe water sources.

We find only weak effects on economic outcomes, suggesting that the interventions did not materially affect households in treated communities (see Appendix Table D.10).

As a robustness check, we control for imbalanced baseline covariates in Appendix Table D.24. Only the effects on community development and political engagement attenuate. Our ANCOVA specification controls for the baseline value of each family, thus addressing the direct effects of any baseline imbalance in that outcome.

Our results prior to the Ebola crisis, under normal conditions, indicate that CM and NFA increased utilization and satisfaction. These effects are not driven by "top-down" improvements in the supply of health services or by greater community contributions to clinics. Rather, they are driven by improvements in the perceived quality of care, which at least partly reflect improvements in the underlying quality of health services. This can be seen in improved patient-provider interactions under NFA, and substantial improvements in child health outcomes under CM.

## 4.2 Longer-run Effects during the Ebola Crisis

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 during the epidemic, including reporting of Ebola cases, as well as mortality among Ebola patients.

#### 4.2.1 Effects on Reporting

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 control sections, while 248 are reported in the treated sections (see Appendix Figure E.1).[28]

[Figure 3 about here.]

We present 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 pre-Ebola findings, where we observe more limited effects of NFA on utilization outcomes, especially in the Ebola sample (see second row of Table 5). Nonetheless, we cannot reject the null hypothesis that CM and NFA have equivalent effects. In the bottom panel of Table 3, we find similar effects using the inverse hyperbolic sine (IHS) transformation of counts. The coefficient on the pooled treatment implies a 41.2 percent increase (see Bellemare and Wichman 2020); the effect of NFA on confirmed cases becomes significant in this specification. We interpret the increase in Ebola cases as reflecting reporting behavior by individuals, rather than effects on transmission: in treatment areas, more individuals reported into clinics to get tested and, if needed, get treatment.

[Table 3 about here.]

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 (based on the pooled treatment effect in the second row of Table 3). We consider the implications of these estimates for the spread of the epidemic. Back-of-the-envelope calculations, following the method employed by Pronyk et al. (2016), suggests that increased reporting by infected individuals reduced the disease's reproduction rate ($R_0$) by 19 percent (see Appendix Section E.10).

[28] The spike in CM observed in March 2015 can be traced to one Section in Bombali district, which registered 28 confirmed cases. We verify that outliers do not drive our estimated effects through leave-one-out and leave-two-out robustness checks (see Appendix Figures E.2 and E.3).

We also observe increases in cases that test negative for the virus. This reinforces our claim (which we substantiate further below) that the effects on total cases reflect increased reporting by individuals seeking testing and treatment.[29]

### 4.2.2 Effects on Patient Deaths

We posit that individuals report more in sections with treated clinics due to improvements in the perceived quality of care. As with the pre-Ebola period, we lack objective measures of the quality of care administered at clinics during the epidemic. However, we are again able to look at a health outcome, which will partly be shaped by the quality of care that was administered—namely, Ebola patient deaths. Sierra Leone lacks vital statistics data, so we can only analyze mortality for cases in the VHF database. 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.

Table 4 presents these results. The first column presents the interaction of caseloads over the last two weeks with the pooled treatment indicator; the second column separates CM and NFA. The pooled results show that patient deaths conditional on reported cases fall disproportionately in treatment areas. When we separate the treatments, we see that these effects are concentrated in CM. For ease of interpretation, Appendix Table E.6 predicts the number of deaths in control and treated sections for a two-week caseload of 2, 5, and 10 cases. We estimate 1 patient death for every 4 cases in control sections; this drops to 1 death for every 7 cases in treated sections—a reduction that is significantly larger in CM, where there is just over 1 patient death for every 10 cases.

[Table 4 about here.]

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 CM sections enjoyed higher survival rates. One may worry that patients in control simply waited longer to report and, thus, presented with greater illness severity and higher risk of mortality. Yet, we show in Appendix Table E.7 that treatment does not reduce the number of days between symptom onset and reporting.

[29] Travel to clinics does not pose a barrier to reporting: the average travel time to clinics in our study is 46 minutes (average travel distance: 3.2 km); national quarantines only lasted a few days at a time.

This fall in Ebola patient deaths, conditional on reported cases, suggests that some factor beyond reporting boosted survival rates. It is consistent with the quality of administered care remaining higher in CM through the crisis period. Indeed, the actions of clinic staff can be highly consequential for Ebola patients' outcomes: effective case management entails vigilantly maintaining hydration; treating symptoms such as high fevers; addressing secondary infections; and calming patients who frequently suffer from acute anxiety (WHO 2016).

Our results during the Ebola crisis parallel those from the pre-crisis period where we observe increased utilization under both treatments, but health outcomes improving under CM alone. Similarly, during the epidemic, we observe increased Ebola reporting under both treatments, but Ebola patient outcomes improving under CM alone. These patterns point to sustained improvements in the perceived quality of care in treatment areas, with larger changes in the effectiveness of care under CM.

#### 4.2.3 Addressing Increased Transmission

We next present evidence to bolster our claim that the interventions did not affect Ebola transmission. The true incidence of Ebola in Sierra Leone is unknown (per Enserink 2014, the WHO and CDC assumed they were missing at least half of all cases). We focus here on plausible channels relating our treatments to transmission and present evidence that such pathways are inoperative.

First, increased transmission could arise from greater interaction among community members in treatment areas. However, meetings associated with the interventions concluded five months before the first Ebola case in Sierra Leone. And, had they continued, meetings are unlikely sites of transmission: Ebola is not an airborne pathogen; rather, it requires direct contact with bodily fluids (e.g., blood, feces, spit, vomit).[30] Those facts notwithstanding, the treatments may still have increased interactions outside the home and, thus, enabled transmission. However, data from contact tracing efforts do not support this possibility. For a subset of infected patients, caseworkers identify people who may have come into contact with the patient. Through this process, they record how contacts are related to the patient (e.g., neighbor, tenant, brother, grandmother). In the last two columns of Appendix Table E.11, we find that contacts outside the nuclear family were, if anything, lower in CM and NFA areas compared to control areas. This pattern is inconsistent with greater interactions and contact outside the family among infected patients in treatment areas.

Second, by increasing the number of individuals reporting into clinics, the treatments could have increased contact between infected and susceptible individuals at these facilities, raising the risk of nosocomial

[30]This is why 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.

transmission (for an account that relates nosocomial transmission to distrust, see Lowes and Montero 2018). 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 positive lab results in the first two days after a patient reports cannot not reflect infections due to exposure after the patient reports into clinics. However, for 92 percent of confirmed cases in our sample, symptom onset occurs prior to reporting, and in 99 percent of cases (all but 2 cases), either symptom onset or lab testing occurs within two days of reporting. This indicates that nearly all confirmed cases we count do not result from infections that occur after the case was reported. (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.)

As further evidence against nosocomial transmission in our sample, Fang et al. (2016) report that infections among health care workers fell precipitously by September 2014 (the start of our Ebola study period), indicating improved awareness and infection control. We continue to find treatment effects in the months after a nationwide effort during November and December 2014 to train health care workers in isolation and no-touch treatment (see Appendix Table E.1 and E.8).

Third, we conduct a placebo test in which we substitute the nearest out-of-sample neighbor for each section. We find no significant effects (see Appendix Table E.9), alleviating concerns that our treated sections are spatially clustered in areas where reporting is higher for reasons unrelated to treatment, such as greater transmission of Ebola.[31]

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. We therefore take a bounding approach, imputing either all ones or all zeros to observations where the ratio is undefined. Imputing all ones assumes that, if cases had been reported, they would have all been confirmed; imputing all zeros assumes that, if cases had been reported, none would have tested positive. Appendix Figure E.5(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.[32]

In Section E.14, we write down a model to clarify what must be assumed for our results to reflect

[31] In Appendix Table E.15, we also look at whether treated sections are more exposed to the epidemic. We find that treated sections are slightly further from index cases in Sierra Leone and Guinea, and that treated sections do not vary systematically in geographic characteristics including road density, the number of rivers, or the ruggedness of terrain.

[32] 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. Appendix Figure E.5(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 Appendix section E.20).

a change in transmission (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 health care needs (Elston et al. 2016).

#### 4.2.4 Addressing Increased Surveillance

Treatment clinics were more exposed to three international NGOs and may have had more communication with the government Health Ministry. This might have increased "top-down" disease surveillance and, thus, increased reported cases.[33]

We first examine rates of contact tracing, which 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 (Appendix Table E.11). Second, we examine three measures derived from the VHF data which also proxy for top-down surveillance efforts: (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. In Appendix Table E.12, we find no significant differences for these variables across treatment and control.

Next, 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.16). There were three types of specialized facilities: Ebola Treatment Units (ETU), Ebola Holding Centers (EHC), and Community Care Centers (CCC). Only one ETU falls within our sample, and it is located in a control section; Appendix Table E.13 shows no significant difference in the counts—either combined or separate—of EHCs or CCCs.[34] Our results in Table 3 are also robust to dropping the small number of sections that contain one or more of these specialized facilities (results available upon request).[35]

[33] We think it is unlikely that our results are instead driven by improved record keeping in treatment clinics. Ebola case investigators who collected the Ebola records were not employees of the clinics, but rather a separate team of surveillance officers hired at the district level. Also, we observe no differential improvements in record keeping between treatment and control clinics in our endline survey (see Appendix Table D.25).

[34] There is only 1 EHC in control sections, 1 in NFA sections, and 2 in CM sections. To address concerns that a small number of sections could drive our results, in Figures E.2 and E.3 we drop all triplets and pairs of triplets as a robustness check.

[35] The presence 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 Appendix Table E.14 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

Finally, it is unlikely that workers at these clinics received more specialized training that would boost their capacity to conduct surveillance as the vast majority of clinic staff nationwide had received training by early December 2014 (see Appendix Table E.1). Unfortunately, only aggregate data is available on the roll-out of training, so we cannot date when individual clinics were covered, but given the pace of the roll-out, all clinics in our sample were likely to have received the training around the same time.

#### 4.2.5 Additional Checks of Ebola Results

We demonstrate robustness to a number of alternative specifications. In Appendix Table E.16, we present estimates using a linear probability model, a Poisson count model, a rare-events logit model, and logged counts (adding one to avoid dropping section-weeks with no cases). In the $\log(y+1)$ transformation, the coefficient on the pooled treatment implies a 45 percent increase, similar in magnitude to the implied effect of 41.2 percent in the IHS specification of Table 3. 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.[36] In Appendix Table E.18, we also estimate a dose-response model which extends the sample to 205 sections, including sections with multiple study clinics, using the proportion of clinics in a section that were treated as the right-hand side variable. We find similar effects under this approach.

To ensure that our results are not driven by a particular place or period, we conduct sub-sample analysis. We re-estimate the pooled effect dropping one matched triplet at a time (Appendix Figure E.2), dropping each possible pair of matched triplets (Appendix Figure E.3), or dropping each week (Appendix Figure E.4). 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 in every month, and we find significant effects in October, December, February and April (Appendix Table E.8). (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 our study period verifies that our estimates are not driven by any particular period. Probable and suspected cases (which constitute 1 and 6.5 percent of total cases, respectively) are included in our count of total reported cases. However, these case types 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 Appendix Table E.19 and find insignificant and negligible treatment effects.[37] In Appendix Table E.20 we

when we use the UNMEER data.

[36] We also collapse the data and estimate cross-sectional models (Appendix Table E.17). Our coefficients are of the same magnitude, but we lose power and precision; the Poisson count models remain highly significant with only 160 observations.

[37] 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.

also subtract probable and suspected cases from total cases and find similar effects. These checks indicate that estimates in Table 3 are driven by increases in the number of patients who report and receive testing.

Next, we address potential imbalance. We aggregate baseline indicators that are unbalanced (see Appendix Table E.4) to the section level and include these as controls. Appendix Table E.21 shows that our results remain unchanged.

We next look for evidence of spillovers between treated and control sections, particularly indications that patients traveled from control to treated sections—reallocation which would amplify our treatment effects on reported cases. Assuming that patients minimize travel costs, spillovers should be largest in treated sections that border (populous) control sections. In Appendix Table E.22, 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. We conduct a number of additional analyses to look for spillovers based on geographic or cultural proximity between treatment and control sections: we look at the distance between clinics (see Appendix Table E.23); connections via road networks (see Appendix Table E.24); and cultural similarity (see Appendix Table E.25). Across these specifications, the point estimates on the interaction terms are all negative, and as a result, the coefficients on the pooled treatment are larger after taking spillovers into account. In addition, imprecision in estimating spillover effects is not a likely source of error in quantifying these spillovers: even when we assume that the true spillover effect is at the "conservative" boundary of its confidence interval and adjust the treatment effect estimates accordingly, they are still large and positive in all cases except one (see Appendix Section E.22).

Finally, confirmed Ebola cases are a relatively rare event, which raises concerns about power. However, note that the standard error on our main result, the pooled treatment effect on total Ebola cases, is 0.084 (Table 3), implying that we would have rejected the null hypothesis of no effect at the 95 percent confidence level for coefficients larger than $0.084 \times 1.96 = 0.165$. Thus, we were powered to detect effects of a magnitude that can reasonably be expected in field studies.

#### 4.2.6 Mechanisms

We interpret the treatment effects on reported Ebola cases to be a consequence of changes in the perceived quality of care provided at CM and NFA clinics.

Concerns about sub-standard care are believed to have 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). If the CM and NFA interventions generated persistent improvements in the perceived quality of health care, 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 Appendix Table D.26 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 effectiveness of Western-style medicine relative to traditional or religious remedies, the primary alternatives to government-run clinics in rural Sierra Leone. (See Appendix Table E.26 for effects on all pre-specified families in the Ebola sample). The effects on general utilization remain positive and significant when we pool the treatments and in CM alone; the effect is attenuated in the NFA arm relative to the full sample. 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 0.15 standard deviations. Finally, we find improvements (about 0.10 standard deviations) in households' attitudes towards Western-style medicine, particularly its effectiveness relative to traditional healers or spiritual remedies. While this indicator 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-style medicine (e.g., Kruk et al. 2015). We combine these three measures into a perceived quality of care index at the household level. In the top row of Table 5, we find that both CM and NFA (pooled and separately) have significant effects on this index.

[Table 5 about here.]

We aggregate this perceived quality of care index to the clinic level, and then instrument it with our treatments to estimate its effect on reported Ebola cases. Table 6 shows a large first-stage effect; the F-statistic of 10.11 satisfies the rule-of-thumb threshold for a strong instrument when we pool the interventions. 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 0.49 cases per section.[40] Appendix Figure E.9 follows Kling, Liebman and Katz (2007) and visualizes this result: we plot the perceived quality of care against total cases (both residualized). The IV estimate can be read off the graph: the (regression-weighted) difference in means between our treatment and control sections along the y-axis corresponds to the reduced form; the (regression-weighted) difference in means along the x-axis

[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 percent of the confirmed Ebola cases.

[39] The satisfaction family 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] Table 6 quantifies the effect of quality of care on reported Ebola cases only under the strong assumption that the entire effect of the instrument works through changes in the perceived quality of care. This approach is similar to Kling, Liebman and Katz (2007) who explore whether the Moving to Opportunity program affects individual outcomes through its effect on neighborhood poverty.

corresponds to the first stage; and the IV coefficient is the ratio of these quantities. We recover near-identical two-stage-least-squares estimates when we use CM and NFA as separate instruments (in the second column of Table 6). The first-stage effect is larger in CM, owing to its larger effect on general utilization in this sub-sample. (The larger first-stage effect for CM aligns with our results in Table 3 where we see larger, if statistically indistinguishable, effects of CM on total cases). These results are consistent with our argument that perceived quality of care becomes critically important in the context of health crises—when fear of sub-standard care might otherwise deter people from seeking testing and treatment.

[Table 6 about here.]

We also find that endogenous changes in the perceived quality of care are associated with greater reporting of total cases in control sections. In Appendix Table E.27, we regress total cases in *control* sections on the change in the perceived quality of care index between baseline and endline. We find a positive relationship, which is significant at the 10% level when we control for population and include fixed effects for week and chiefdom, the administrative unit just above sections. The magnitude of the correlational relationship is about half as large as that estimated using two-stage least squares, suggesting that it understates the effect of quality of care on reporting of Ebola cases.

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 organization and services (see Appendix Table E.26).[41]

## 4.3 Cost Effectiveness

Our results show that the CM and NFA interventions not only improve outcomes under normal conditions, but also facilitate the reporting of Ebola cases, helping to contain the spread of the epidemic. Bolstering the health system's resilience—its capacity to mount an effective response to such a crisis—is an unintended consequence of these programs, which were established to improve maternal and child health care during normal times. Putting aside their intended purpose (and benefits during business-as-usual periods), we ask whether these interventions constitute cost-effective approaches to containing epidemics like Ebola. In short, do their effects on containment alone justify spending on these programs in advance of an epidemic?

We pit these interventions against a well-regarded but reactive approach to Sierra Leone's Ebola crisis. Community Care Centers (CCCs) were set up after the Ebola crisis hit to allay fears about Western-style

[41] Separating the two treatments, we observe clinic organization and services increases in NFA. We also observe contributions to clinics increase under CM: community members report additional contributions of time, money, and/or labor to clinics; clinic staff do not, however, report a significant increase in contributions to the clinic (results available upon request).

medical facilities and, thus, encourage reporting and early isolation and treatment (Michaels-Strasser et al. 2015). They were widely considered effective and low cost among the emergency response centers (as compared to the Ebola Holding Cells and Ebola Treatment Units, which both provided intensive care and treatment). For example, CCCs were typically set up in tents or re-purposed buildings, and did not require new construction.

The cost of a CCC was $707,274 on average. In a quasi-experimental evaluation of CCCs, we find that these centers were indeed successful in encouraging Ebola reporting (Christensen et al. 2020). Specifically, we find sections with CCCs saw 0.54 additional cases tested per section-week, of which 0.129 were confirmed to be Ebola (Christensen et al. 2020, Table 1). Over the full crisis period (34 weeks), this amounted to 18.50 reports and 4.39 confirmed cases per section.

In contrast, the pooled CM/NFA intervention led to 0.173 additional cases tested per section-week, of which 0.059 were confirmed to be Ebola cases. Over the crisis, this totaled 5.88 reports and 2 confirmed cases per section. The cost of the pooled CM/NFA intervention is $6,375 per clinic (see Appendix Section E.26 for details). Comparing the estimated effect size to the cost for each intervention shows that CCCs increased testing at a cost of $38,232 per case. In comparison, the pooled intervention cost only $1,084 per case. For confirmed cases, the numbers are $161,115 and $3,188, respectively.

Whether the pooled interventions or the emergency CCCs are more cost-effective in managing epidemic outcomes depends on the likelihood of an epidemic such as Ebola breaking out. Absent an epidemic, no money is spent on reactive measures, like CCCs; and the interventions incur costs without contributing to containment. Comparing the ratios of cost and effect sizes implies that the interventions are more cost-effective than CCCs for epidemic events with $>$2–3 percent probability of occurring (see Appendix Section E.26).[42]

Simulations based on historical data suggest that the annualized likelihood of an epidemic of comparable magnitude to the 2014–15 Ebola outbreak is similar (Stephenson et al. 2020). This suggests that preemptive investments in public health, similar to our CM and NFA treatments, are worth making—not just for their immediate effects on community health, but as cost-effective ways of building resilience to future outbreaks.

[42]This statement again focuses solely on epidemic outcomes. While CM and NFA influence a wider range of outcomes (e.g., clinic utilization and child health), we refrain from comparing their cost-effectiveness against CCCs using pre-Ebola measures for two reasons. First, CCCs were established after our endline surveys—which means we lack common pre-Ebola indicators for this comparison. Moreover, since CCCs are not designed to be a business-as-usual intervention and would not exist except under epidemic conditions, measuring their cost-effectiveness in altering these other outcomes is also less relevant.

# 5. Conclusion

We use a randomized experiment completed less than a year before the Ebola outbreak in Sierra Leone to test the effectiveness of two interventions that harness social incentives and promote accountability: one implemented community monitoring of government-run health clinics, and the other conferred status awards to clinic staff. Our findings suggest that these interventions 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.

In the period prior to the Ebola crisis, both interventions increase patient satisfaction and clinic utilization, a revealed preference measure that reflects individuals' perceptions about the quality of care provided. The community monitoring intervention also dramatically reduced under-5 mortality, suggesting that improvements in perceived quality at least partly reflect provider behavior and changes in the actual quality of care delivered at these clinics.

We evaluate these programs' longer-run effects during the Ebola crisis. Ebola containment requires early isolation. Yet, concerns about sub-standard health care and a lack of confidence in health workers deterred patients in Sierra Leone from reporting to clinics.

We find that both interventions substantially increased reporting of Ebola cases, by about 60%. In addition, analogous to the pre-Ebola period, CM also improved health outcomes during the crisis, reducing Ebola patient mortality conditional on cases. These results suggest that improvements in the perceived quality of care at intervention clinics led to increased reporting during the crisis, and improvements in administered care in CM clinics also persisted into the crisis period. CM has qualitatively stronger effects than NFA both before the crisis, and during the Ebola outbreak; this suggests that involving the community in promoting accountability may be especially effective in improving the quality of health services. One possible reason for this effect is that public meetings act as coordination devices where community members can align beliefs and perceptions about the clinic.

We find no support for two alternative explanations of why the interventions could have increased reported case counts—by unintentionally increasing transmission or enabling more top-down surveillance. Inconsistent with increased transmission at treated clinics, we observe increased reporting by both individuals who tested positive and those who tested negative, with no observed changes in the ratio of positive to negative case types. 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.

Together, our results suggest that these interventions have the power not just to improve health systems over the short run, but also to boost their resilience to crises that emerge over the longer run. While the increases in patient utilization in the pre-Ebola period are modest, the effects on reporting during the Ebola epidemic are substantial. This suggests that even moderate shifts in the perceived quality of care can strengthen health systems during crises, and pay substantial dividends during these critical periods. Because such effects are difficult to capture, it remains an important open question whether these types of interventions bolster reporting and resiliency in other places and crises. Our analysis of mechanisms suggests that such effects should especially manifest themselves where the baseline (perceived) quality of local health care is low—a condition that Kruk et al. (2018) find is all too common across low- and middle-income countries. If these interventions are also effective in other settings, they could constitute a promising approach to preparing for future crises.

# References


Abramowitz, Sharon, Braeden Rogers, Liya Akilu, Sylvia Lee, and David Hipgrave (2016). "Ebola Community Care Centers: Lessons learned from UNICEF's 2014-2015 Experience in Sierra Leone." Technical report.

Alsan, Marcella and Marianne Wanamaker (2017). "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, Edward Davenport, and Scott S. Lee (2020). "Losing prosociality in the quest for talent? sorting, selection, and productivity in the delivery of public services." *American Economic Review* 110, 1355–1394.

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.

Ball, Sheryl, Catherine Eckel, J Grossman, Philip, and William Zame (2001). "Status in markets." *Quarterly Journal of Economics* 111(1), 161–188.

Banerjee, Abhijit, Angus Deaton and Esther Duflo (2004). "Health care delivery in rural Rajasthan." *Economic and Political Weekly* 39(9), 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." *Working paper, Georgetown University*.

Bellemare, Marc F. and Casey J. Wichman (2020). "Elasticities and the inverse hyperbolic sine transformation." *Oxford Bulletin of Economics and Statistics* 82(1), 50–61.

Bénabou, 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). "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). "Experimental Evidence on the Long-Run Impact of Community-Based Monitoring." *American Economic Journal: Applied Economics* 9(1), 33–69.

Björkman Nyqvist, Martina, Andrea Guariso, Jakob Svensson, and David Yanagizawa-Drott (2019). "Reducing Child Mortality in the Last Mile: Experimental Evidence on Community Health Promoters in Uganda." *American Economic Journal: Applied Economics* 11(3), 155–92.

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.

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 (2020). "Community-based Crisis Response: Evidence from Sierra Leone's Ebola Outbreak." *AEA Papers and Proceedings* 110, 260–64.

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.

Deaton, Angus (2013). *The Great Escape: Health, Wealth, and the Origins of Inequality*. Princeton University Press.

Denney, Lisa and Richard Mallett (2014). "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). "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). "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.

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

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). "Variability in Intrahousehold Transmission of Ebola Virus, and Estimation of the Household Secondary Attack Rate." *The Journal of Infectious Diseases* 217(2), 232–237.

Greevy, Robert and Cole Beck (2016). "nbpMatching Demo: Triplet Matching Prior to Randomization." (http://biostat.mc.vanderbilt.edu/wiki/Main/MatchingTripletsPriorToRandomization).

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

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

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

Kosfeld, Michael and Susanne Neckermann (2011). "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). "What is a resilient health system? Lessons from Ebola." *The Lancet* 385(9980), 1910–1912.

Leone, Statistics Sierra, University of Sierra Leone, Catholic Relief Services, and ICF (2016). "Sierra leone malaria indicator survey 2016. https://www.afro.who.int/sites/default/files/2017-05/slmis.pdf, accessed on October 12, 2020.

Levy, Benjamin, Carol Y Rao, Laura Miller, Ngozi Kennedy, Monica Adams, Rosemary Davis, Laura Hastings, Augustin Kabano, Sarah D Bennett, and Momodu Sesay (2015). "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).

Makori, Christine (2012). "Official documents - agreement for dsdp grant tf012665 (english)." Technical Report http://documents.worldbank.org/curated/en/651161468333621328/Official-Documents-Agreement-for-DSDP-Grant-TF012665, World Bank.

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). "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.

MOH (2017). "Human resources for health strategy 2017-2021. https://www.afro.who.int/sites/default/files/2017-05/hrhstrategy2017.pdf, last accessed on July 13, 2020.

Mohanan, Manoj, Vikram S Rajan, Kendal Swanson, and Harsha Thirumurthy (2020). "Information and Facilitation Interventions for Accountability in Health and Nutrition: Evidence from a Randomized Trial in India." *Economic Research Initiatives at Duke (ERID) Working Paper* (295).

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.

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.

Owada, Kei, Tim Eckmanns, Kande-Bure O'Bai Kamara, and Olushayo Oluseun Olu (2016). "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.

Özler, Berk, Lia C.H. Fernald, Patricia Kariger, Christin McConnell, Michelle Neuman, and Eduardo Fraga (2018). "Combining pre-school teacher training with parenting education: A cluster-randomized controlled trial." *Journal of Development Economics* 133, 448–467.

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). "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*.

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

Stephenson, N., K. Miller, M. Gallivan, C. Lam, V. Serhiyenko, and N. Madhav (2019). "Risk management and preparedness: Use of stochastic modeling and risk analytics to estimate frequency and severity of filovirus epidemics." *International Journal of Infectious Diseases* 79, 125.

Stephenson, N., K. Miller, M. Gallivan, C. Lam, V. Serhiyenko, and N. Madhav (2020). "Filovirus Model Catalog v2.

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). "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). "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.

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/).

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

WHO (2014b). "Key considerations for the implementation of community care centres." Technical Report https://www.who.int/csr/resources/publications/ebola/community-care-centres/en/, World Health Organization.

WHO (2016). "Clinical management of patients with viral haemorrhagic fever." Technical Report https://www.who.int/csr/resources/publications/clinical-management-patients/en/, WHO.

**Figure 1:** Consort Diagram

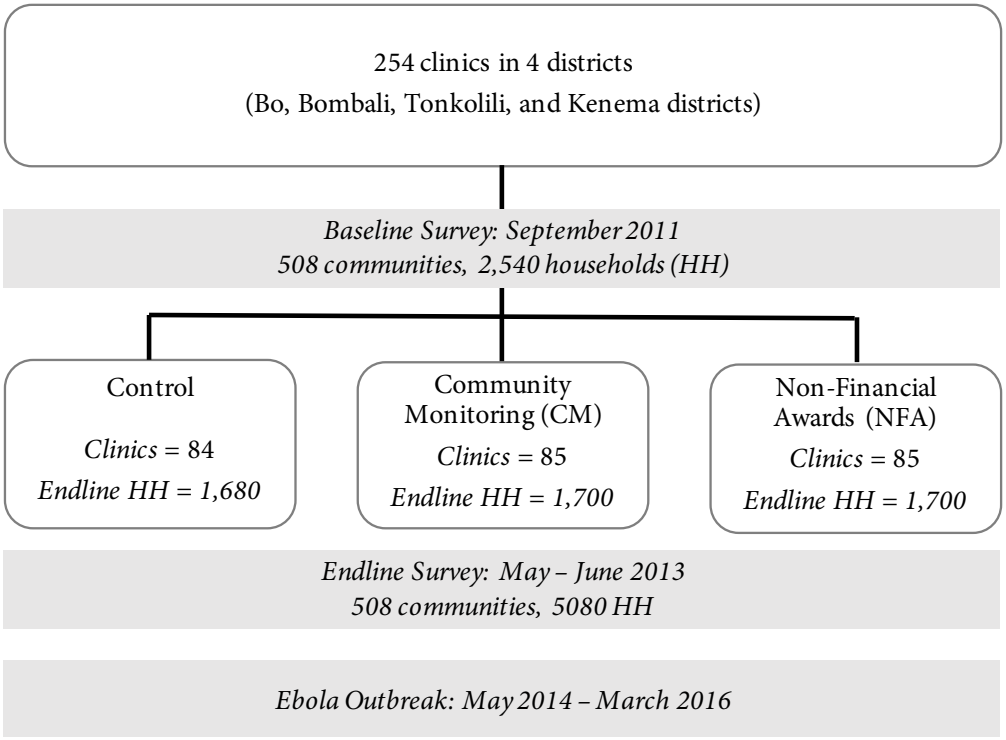
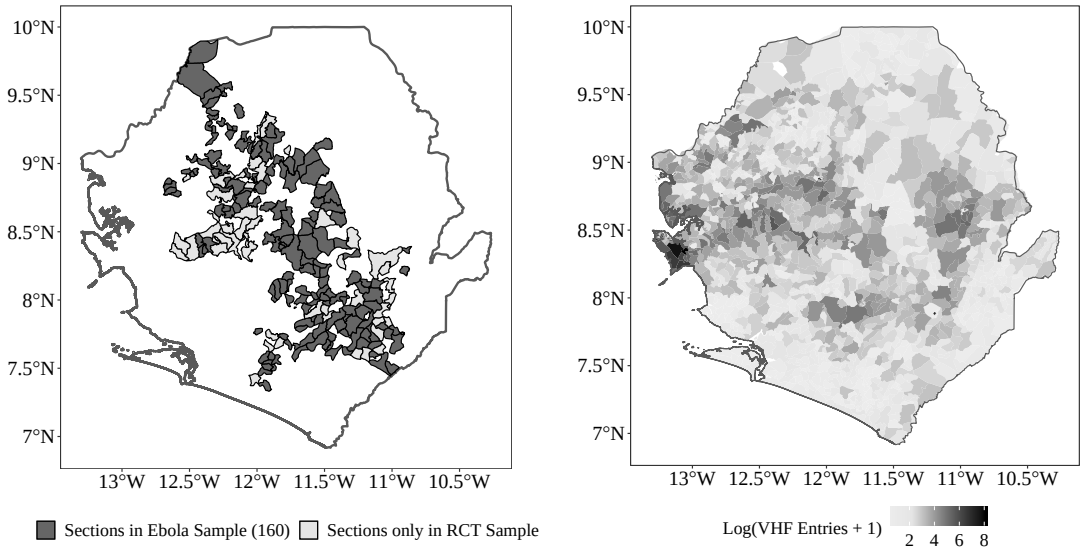



Figure 1: Samples and timing associated with baseline and endline surveys, randomization, and Ebola crisis. The crisis was initially declared over in November 2015; however, a few additional cases subsequently emerged, and the country was finally deemed "Ebola free" in March 2016.

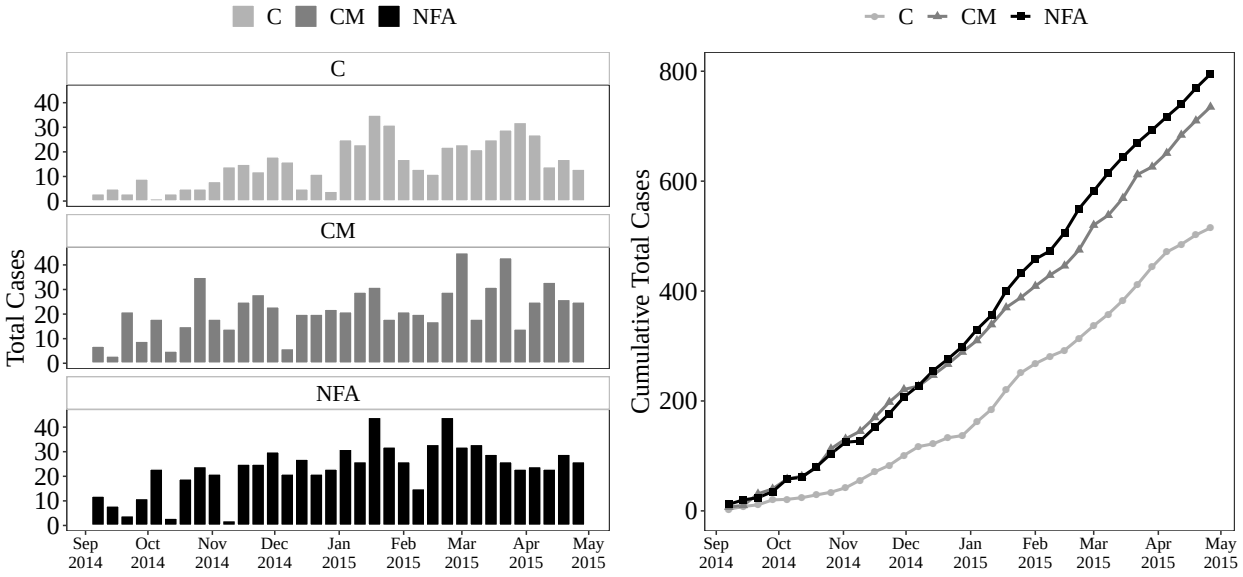
**Figure 2:** Mapping of Ebola Cases and Sample




**(a)** Sections Used in Ebola Analysis **(b)** Heat Map of VHF Cases by Section

Figure 2(a): Map of all sections that contain clinics that were part of the original randomized experiment. The 45 sections in light gray are excluded from the primary Ebola analysis, because they contain more than one clinic from the original RCT. Figure 2(b): The number of entries by section in the Viral Hemorrhagic Fever (VHF) database maintained by the Sierra Leone Ministry of Health with support from the CDC during the Ebola crisis. We log the counts, first adding one to avoid dropping sections with no entries.

**Figure 3:** Total Ebola Cases by Treatment




**(a)** Weekly Counts

**(b)** Cumulative Counts

Figure 3(a) plots the time series of total Ebola cases by week; bars represent the raw counts. C refers to control (54 sections); CM refers to community monitoring (46 sections); NFA refers to non-financial awards (60 sections). We use the date that the case was first saved in the VHF. Figure 3(b) graphs the cumulative count of total Ebola cases by treatment group.

**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 | 0.112 | 0.126 | 0.099 | 0.026 | 7.054 | 4496 |
| | (1.000) | (0.031)*** | (0.034)*** | (0.037)*** | (0.033) | (0.001)*** | |
| | | [0.005]*** | [0.003]*** | [0.032]** | | | |
| Maternal utilization | 0.000 | 0.061 | 0.175 | −0.043 | 0.218 | 4.128 | 888 |
| | (1.000) | (0.064) | (0.077)** | (0.076) | (0.081)*** | (0.017)** | |
| | | [0.327] | [0.068]* | [0.548] | | | |
| Satisfaction | 0.000 | 0.097 | 0.086 | 0.108 | −0.022 | 2.840 | 5052 |
| | (1.000) | (0.041)** | (0.047)* | (0.048)** | (0.048) | (0.060)* | |
| | | [0.038]** | [0.095]* | [0.048]** | | | |
| Health outcomes | 0.000 | 0.064 | 0.166 | −0.039 | 0.205 | 8.105 | 5053 |
| | (1.000) | (0.051) | (0.055)*** | (0.060) | (0.055)*** | (0.000)*** | |
| | | [0.265] | [0.014]** | [0.548] | | | |

*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. Column 1 reports standard deviation in parentheses. Columns 2–5 report robust standard errors, clustered by clinic. Multiple-inference corrected $q$-values that adjust for the false discovery rate within treatment arm across all ten pre-specified outcomes are shown in square brackets. The $F$-test column provides evidence on 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.

**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 |
|---|---|---|---|---|---|---|---|
| **Supply Side Measures and Community Support** | | | | | | | |
| Health service delivery | 0.000 | 0.026 | 0.041 | 0.022 | 0.019 | 0.141 | 2877 |
| | (1.000) | (0.057) | (0.079) | (0.060) | (0.071) | (0.869) | |
| | | [0.395] | [0.266] | [0.583] | | | |
| Clinic organization and services | 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] | | | |
| Community support | 0.000 | 0.033 | 0.044 | 0.021 | 0.023 | 0.079 | 508 |
| | (1.000) | (0.095) | (0.112) | (0.109) | (0.113) | (0.924) | |
| | | [0.537] | [0.504] | [0.619] | | | |
| **Community development projects and infrastructure** | | | | | | | |
| 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]** | | | |
| Water and sanitation | 0.000 | 0.156 | 0.093 | 0.219 | −0.126 | 4.566 | 5053 |
| | (1.000) | (0.063)** | (0.073) | (0.073)*** | (0.072)* | (0.011)** | |
| | | [0.038]** | [0.202] | [0.030]** | | | |

*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. Column 1 reports standard deviation in parentheses. Columns 2–5 report robust standard errors, clustered by clinic. Multiple-inference corrected $q$-values that adjust for the false discovery rate within treatment arm across all ten pre-specified outcomes are shown in square brackets. The $F$-test column provides evidence on 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.

**Table 3:** Reported Ebola Cases per Section per Week

| | (1) Control Mean | (2) Pooled | (3) CM | (4) NFA | (5) Difference | (6) 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 | 0.206 | 0.083 | 0.096 | 0.074 | 0.022 | 5,440 |
| | (0.47) | (0.043)* | (0.057)* | (0.051) | (0.065) | |
| Confirmed | 0.009 | 0.029 | 0.035 | 0.025 | 0.01 | 5,440 |
| | (0.1) | (0.01)*** | (0.015)** | (0.012)** | (0.017) | |
| Negative | 0.179 | 0.058 | 0.052 | 0.063 | -0.011 | 5,440 |
| | (0.433) | (0.035)* | (0.045) | (0.043) | (0.052) | |

*Notes*: Treatment effects estimated using OLS including matching-triplet and week fixed effects. Column 1 reports standard deviation in parentheses. Columns 2–4 report robust standard errors, clustered by section, in parentheses. The 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 ($\text{IHS}(y) = \log(y + \sqrt{(1+y^2)})$). Total cases include confirmed, negative, suspected and probable cases. Significance: * is significant at the 10% level, ** is significant at the 5% level and *** is significant at the 1% level.

**Table 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 × Pooled | −0.098 | |
| | $(0.043)^{**}$ | |
| CM | | 0.116 |
| | | $(0.037)^{***}$ |
| Total Cases in Last 2 Weeks × CM | | −0.149 |
| | | $(0.046)^{***}$ |
| NFA | | −0.007 |
| | | (0.025) |
| Total Cases in Last 2 Weeks × NFA | | −0.019 |
| | | (0.032) |
| Control Mean | 0.149 | 0.149 |
| | (0.49) | (0.49) |
| Observations | 5,280 | 5,280 |

*Notes*: Treatment effects estimated using OLS including matching-triplet and week fixed effects. Robust standard errors, clustered by section, in parentheses. Significance: * is significant at the 10% level, ** is significant at the 5% level and *** is significant at the 1% level.

**Table 5:** Perceived Quality of Care

| | (1) Control Mean | (2) Pooled | (3) CM | (4) NFA | (5) Difference | (6) Joint $F$-test ($p$) | (7) N |
|---|---|---|---|---|---|---|---|
| **Perceived quality of care** | 0.000 | 0.167 | 0.181 | 0.156 | 0.025 | 6.069 | 3173 |
| | (1.000) | (0.048)*** | (0.062)*** | (0.053)*** | (0.063) | (0.003)*** | |
| General utilization | 0.000 | 0.098 | 0.135 | 0.070 | 0.065 | 2.792 | 2085 |
| | (1.000) | (0.045)** | (0.058)** | (0.047) | (0.053) | (0.064)* | |
| Satisfaction with public health workers | 0.000 | 0.151 | 0.150 | 0.152 | −0.002 | 4.807 | 3149 |
| | (1.000) | (0.049)*** | (0.065)** | (0.053)*** | (0.064) | (0.009)*** | |
| Relative effectiveness of western medicine | 0.000 | 0.102 | 0.117 | 0.090 | 0.026 | 2.040 | 2663 |
| | (1.000) | (0.051)** | (0.064)* | (0.054)* | (0.060) | (0.133) | |

*Notes:* Treatment effects are estimated using Missing-Indicator ANCOVA, controlling for the community-level average of the outcome at baseline and matching-triplet fixed effects. Column 1 reports standard deviation in parentheses. Columns 2–5 report robust standard errors, clustered by clinic. The $F$-test column provides evidence on the joint significance of CM and NFA, with the associated $p$-value in parentheses. "Perceived Quality of Care" is an equally weighted index of the other three variables in the table. Significance: * is significant at the 10% level, ** is significant at the 5% level and *** is significant at the 1% level.

**Table 6:** Perceived Quality of Care and Ebola Cases

| | (1) | (2) |
|---|---|---|
| First-stage | Perceived Quality of Care | |
| Pooled (CM or NFA) | 0.394 | |
| | (0.124)*** | |
| CM | | 0.480 |
| | | (0.156)*** |
| NFA | | 0.329 |
| | | (0.137)** |
| Two-stage Least Squares | Total Cases | |
| Perceived Quality of Care | 0.494* | 0.491* |
| | (0.254) | (0.255) |
| First-stage F-statistic | 10.11 | 5.36 |
| Observations | 5,440 | 5,440 |

*Notes*: Models include triplet and week fixed effects, as well as the baseline value of the perceived quality of care index. Robust standard errors, clustered by section, in parentheses. "Perceived Quality of Care" is the index shown in Table 5. Significance: * is significant at the 10% level, ** is significant at the 5% level and *** is significant at the 1% level. In column (2), the $J$ statistic is 0.006 with an associated p-value of 0.94.

# Online Appendix

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

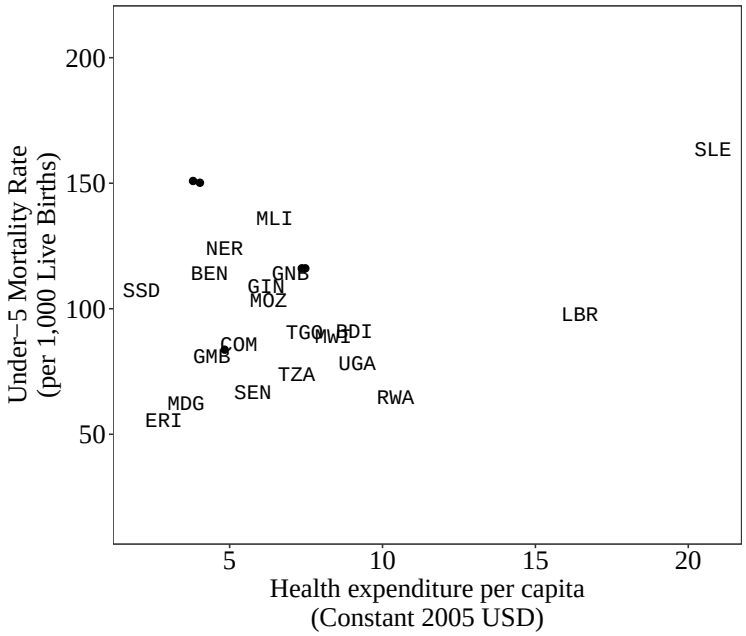
Darin Christensen (UCLA)
Oeindrila Dube (University of Chicago)
Johannes Haushofer (Stockholm University)
Bilal Siddiqi (UC Berkeley, CEGA)
Maarten Voors (Wageningen University)

# Contents

# A. Context

## A.1 Cross-National Health Indicators

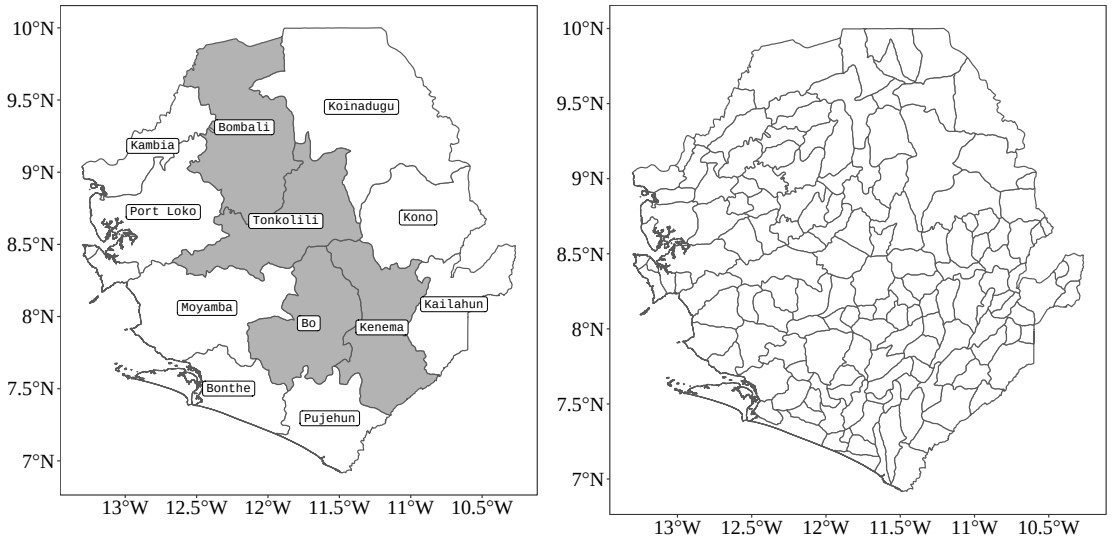
**Figure A.1:** Health Expenditure and Under-5 Mortality in 2010




We use data 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. We omit some country codes to avoid over-plotting; those observations appear as dots.

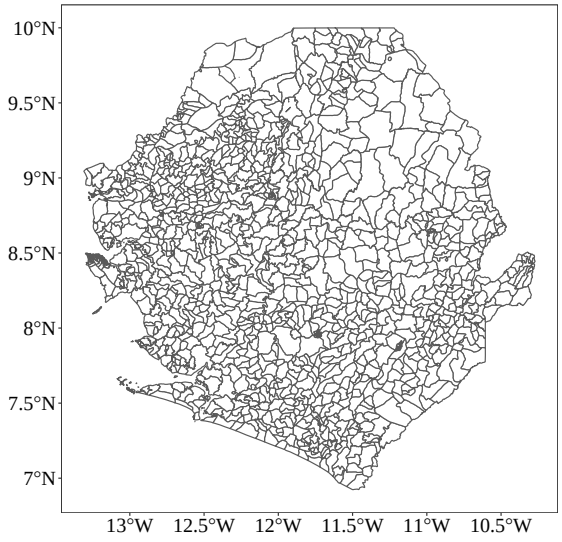
## A.2 Administrative Boundaries

**Figure A.2:** Administrative Boundaries




**(a)** Districts

**(b)** Chiefdoms




**(c)** Sections

Maps of Sierra Leone's different administrative units are provided for reference. Sections (Appendix Figure A.2(c)) nest neatly in chiefdoms (Appendix Figure A.2(b), which nest neatly in districts (Appendix Figure A.2(a)). The randomized experiment was run in four districts: Bombali, Tonkolili, Bo, and Kenema.

# B. Variable descriptions

## B.1 Pre-specified outcome variables and deviations

Below we detail the families of medium-term outcome variables, which 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. Deviations from our analysis plan (AP) are detailed using footnotes.

1. **General Utilization Index**[A1]
   - (a) Health episodes in response to which individuals visited clinic {i}
   - (b) Use of traditional healers {i}[A2]
2. **Maternal Utilization Index (maternal episodes: among mothers who have given birth in the last year)**
   - (a) Antenatal/postnatal care index [standardized summary index of i–ii] {i}
     - i. Number of ANC visits
     - ii. Number of PNC visits
   - (b) Childbirth in facility {i}
     - i. Proportion of pregnant mothers who gave birth in facility
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}
   - (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) Vaccines: (Among households with eligible children) [standardized 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)

[A1] The AP specified examining visits to seek care at clinics as a proportion of health episodes. We use number of visits as the preferred measure since it captures both changes in the propensity to report a health episode, and the propensity to seek care at a clinic conditional on having reported an episode. We draw this distinction since the treatment may have altered what individuals conceptualize to be a health need or health episode. However, all our results are qualitatively similar if we instead use either the proportion of visits or an indicator of any visits to a Western-style clinic. These results are available upon request.

[A2] The AP 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 were 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 in our main specifications. In addition, we also conduct a robustness check with both the utilization of Western clinics and traditional healers, restricting to the illness/injury episodes where both variables can be measured, in Appendix Table D.22.

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 Organization and Services**

(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**[A3]

(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 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

[A3]Per our AP, two satisfaction outcomes appear in both this family and the satisfaction family (measured at the individual and household level, respectively). We verify that this redundancy does not affect the health service delivery index result.

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) 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.)[A4]

(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)**[A5]

(a) Development projects (Excluding NGOs) {com}[A6]

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)

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?

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**

[A4]The AP mis-specified financial contributions as originating from the community survey, instead of the clinic survey.

[A5]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 gauge the monitoring mechanism more directly. However, we verify that omitting these variables from the CDPE index has no consequence on the estimated effect (available upon request).

[A6]In the ANCOVA specification, we control for projects in the past two years from the baseline survey.

(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) Bicycle, l) Motorcycle or Motor scooter, m) Animal-drawn cart, n) Car or Truck, o) Boat with no Motor, p) 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.
    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.

(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?[A7]

[A7] The AP also included retails prices, however, these were not properly recorded and are therefore omitted from the index.

## B.2 Descriptive statistics

**Table B.1:** Descriptive Statistics at Endline

| | Median | Mean | SD | Min | Max | N | BL data |
|---|---|---|---|---|---|---|---|
| *General utilization* | | | | | | | |
| (1) Number of health episodes in which sought Western care | 1 | 0.988 | 0.382 | 0 | 3 | 4496 | Yes |
| *Maternal utilization* | | | | | | | |
| (1) ANC/PNC visits index | −0.187 | −0.054 | 0.956 | −1.779 | 5.602 | 887 | Yes |
| (2) Birth in Western style facility | 1 | 0.862 | 0.345 | 0 | 1 | 877 | Yes |
| *Health outcomes* | | | | | | | |
| (1) U5 Child death per HH | 0 | 0.033 | 0.178 | 0 | 1 | 5053 | Yes |
| (2) Maternal death per HH | 0 | 0.001 | 0.034 | 0 | 1 | 5053 | Yes |
| (3) Illness/injury in HH | 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 | −0.619 | −0.002 | 0.975 | −0.619 | 2.649 | 856 | Yes |
| (7) Child illness index | −0.091 | 0.028 | 0.924 | −1.435 | 3.424 | 4993 | Yes |
| *Satisfaction* | | | | | | | |
| (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 | 4 | 3.658 | 0.670 | 1 | 4 | 2535 | Yes |
| (4) Individual would return to clinic | 1 | 0.969 | 0.168 | 0 | 1 | 2527 | Yes |
| *Health service delivery* | | | | | | | |
| (1) Absenteeism index | −0.239 | 0.057 | 1.070 | −0.663 | 7.471 | 2874 | Yes |
| (2) Paid for treatment | 0 | 0.404 | 0.491 | 0 | 1 | 2872 | Yes |
| (3) Amount paid | 0 | 7816.945 | 35359.126 | 0 | 1360000 | 2843 | Yes |
| (4) Any problem | 0 | 0.061 | 0.240 | 0 | 1 | 2869 | Yes |
| (5) Staff not present | 0 | 0.020 | 0.140 | 0 | 1 | 2869 | Yes |
| (6) Drugs not available | 0 | 0.027 | 0.162 | 0 | 1 | 2869 | Yes |
| (7) Facility not clean | 0 | 0.002 | 0.042 | 0 | 1 | 2869 | Yes |
| (8) Unpleasant staff behavior | 0 | 0.021 | 0.142 | 0 | 1 | 2869 | Yes |
| (9) Medicine always in stock | 1 | 0.948 | 0.221 | 0 | 1 | 2478 | No |
| (10) Individual satisfaction with care | 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 | 4 | 3.752 | 0.555 | 1 | 4 | 2845 | Yes |
| *Clinic organization and services* | | | | | | | |
| (1) Clinic service provision index | 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 |
| *CDPE* | | | | | | | |
| (1) Projects with local council/chief | 0 | 0.089 | 0.285 | 0 | 1 | 507 | Yes |
| (2) Community provided labor | 0 | 0.073 | 0.261 | 0 | 1 | 504 | Yes |
| (3) Community involved in planning | 0 | 0.050 | 0.217 | 0 | 1 | 504 | Yes |
| (4) Problem addressed collectively? | 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 |
| *Contributions to clinic* | | | | | | | |
| (1) Contributions to clinic index (community survey) | 0.083 | 0.075 | 0.934 | −1.170 | 4.193 | 508 | Yes |
| (2) Contributions to clinic index (facility survey) | 0.155 | −0.034 | 1.065 | −4.528 | 3.331 | 508 | Yes |
| *Water and sanitation* | | | | | | | |
| (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 |
| *Economic outcomes* | | | | | | | |
| (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 | −0.225 | 0.016 | 1.031 | −1.311 | 14.540 | 5053 | Yes |

# C. Manipulation Check and Balance

## C.1 Manipulation Checks

**Table C.1:** Manipulation Checks: Community Meetings

| | (1) Control Mean | (2) Pooled | (3) CM | (4) NFA | (5) Difference | (6) N |
|---|---|---|---|---|---|---|
| (1) A community meeting by PLAN, Concern, or IRC took place | 0.435 | 0.269 | 0.421 | 0.115 | 0.306 | 506 |
| | (0.497) | (0.047)*** | (0.050)*** | (0.053)** | (0.049)*** | |
| (2) How many meetings took place? | 0.897 | 1.055 | 1.631 | 0.483 | 1.147 | 498 |
| | (1.218) | (0.142)*** | (0.156)*** | (0.157)*** | (0.164)*** | |
| (3) Was village informed of meeting outcomes? | 0.411 | 0.266 | 0.403 | 0.127 | 0.276 | 506 |
| | (0.493) | (0.049)*** | (0.054)*** | (0.055)** | (0.053)*** | |

*Notes:* Treatment effects are estimated using cross-sectional OLS at endline and including matching-triplet fixed effects. Column 1 reports standard deviation in parentheses. Columns 2–5 report robust standard errors, clustered by clinic. 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: Competitions

| | (1) Control Mean | (2) Pooled | (3) CM | (4) NFA | (5) Difference | (6) N |
|---|---|---|---|---|---|---|
| (1) Clinic staff heard of a competition? | 0.476 | 0.331 | 0.190 | 0.473 | −0.282 | 254 |
| | (0.502) | (0.063)*** | (0.073)*** | (0.064)*** | (0.059)*** | |
| (2) Clinic participated in a competition? | 0.155 | 0.428 | 0.198 | 0.657 | −0.459 | 254 |
| | (0.364) | (0.061)*** | (0.067)*** | (0.063)*** | (0.064)*** | |

*Notes:* Treatment effects are estimated using cross-sectional OLS at endline and including matching-triplet fixed effects. Column 1 reports standard deviation in parentheses. Columns 2–5 report robust standard errors. Significance: * is significant at the 10% level, ** is significant at the 5% level and *** is significant at the 1% level.

## C.2 Balance

**Table C.3:** Baseline Balance

| | (1) Control Mean | (2) CM–Control Difference | (3) NFA–Control Difference | (4) CM–NFA Difference | (5) N |
|---|---|---|---|---|---|
| *Village characteristics* | | | | | |
| Motorable road | 0.890 | −0.009 | 0.005 | −0.014 | 503 |
| | (0.314) | (0.036) | (0.035) | (0.035) | |
| Mobile phone coverage within 1 mile from the community | 0.798 | 0.058 | 0.096 | −0.038 | 504 |
| | (0.402) | (0.044) | (0.041)** | (0.037) | |
| Distance to the closest clinic | 1.257 | −0.204 | 0.338 | −0.542 | 504 |
| | (1.868) | (0.329) | (0.481) | (0.463) | |
| Travel cost to closest clinic | 73.529 | −24.273 | −24.389 | 0.116 | 503 |
| | (740.324) | (72.811) | (74.502) | (65.453) | |
| *Household characteristics and questions to household head* | | | | | |
| Household size | 4.452 | −0.061 | 0.007 | −0.068 | 4774 |
| | (3.068) | (0.056) | (0.058) | (0.059) | |
| Number of illness or injury cases per household | 0.071 | −0.039 | −0.026 | −0.013 | 4774 |
| | (0.270) | (0.011)*** | (0.012)** | (0.013) | |
| Birth in household last year | 0.158 | −0.028 | 0.009 | −0.037 | 2127 |
| | (0.365) | (0.014)** | (0.015) | (0.014)** | |
| Child under 2 in household | 0.234 | −0.013 | 0.027 | −0.040 | 2126 |
| | (0.424) | (0.018) | (0.018) | (0.018)** | |
| Prominent village member in household | 0.041 | −0.007 | −0.002 | −0.005 | 2090 |
| | (0.199) | (0.010) | (0.010) | (0.009) | |
| Believes doctor's advice | 0.994 | 0.000 | −0.007 | 0.007 | 1977 |
| | (0.078) | (0.004) | (0.005) | (0.004)* | |
| Health care fees unaffordable | 2.317 | 0.023 | 0.030 | −0.007 | 2057 |
| | (0.798) | (0.045) | (0.050) | (0.046) | |
| Trust in the community | 1.848 | −0.032 | −0.010 | −0.021 | 2127 |
| | (0.667) | (0.051) | (0.049) | (0.052) | |
| Community cohesion | 2.421 | −0.018 | 0.009 | −0.026 | 2122 |
| | (0.604) | (0.034) | (0.036) | (0.035) | |
| Believe VHC members represent your interest | 2.761 | 0.094 | 0.159 | −0.066 | 984 |
| | (1.084) | (0.107) | (0.122) | (0.104) | |
| The VHC can be trusted | 2.443 | −0.171 | −0.089 | −0.083 | 1148 |
| | (0.963) | (0.103)* | (0.106) | (0.101) | |
| *Individual characteristics* | | | | | |
| Muslim | 0.849 | −0.037 | −0.017 | −0.021 | 9761 |
| | (0.358) | (0.026) | (0.026) | (0.029) | |
| Mende (Ethnicity) | 0.432 | −0.019 | −0.005 | −0.014 | 9759 |
| | (0.495) | (0.012) | (0.012) | (0.011) | |
| Temne (Ethnicity) | 0.333 | 0.023 | 0.078 | −0.055 | 9759 |
| | (0.471) | (0.039) | (0.036)** | (0.035) | |
| Highest level of education | 1.808 | 0.090 | 0.323 | −0.233 | 9379 |
| | (2.923) | (0.103) | (0.109)*** | (0.112)** | |
| *Chi*-square *p*-value | | 0.564 | 0.115 | 0.584 | |

*Notes:* This table presents baseline balance for the sample of households included in both baseline and endline surveys. Column (1) shows the mean and standard deviation of the control group at baseline. Columns (2) and (3) indicate the regression coefficients and standard errors on the CM and NFA treatment arm indicators compared to the control, with standard errors clustered at the clinic level in parentheses. Column (4) compares the CM treatment arm to the NFA arm. Column (5) displays the sample size. The top panel reports village level data. The village sample is comprised of two villages from the catchment areas of 254 clinics (resulting in a sample of 508 villages, with some missing data). The second panel reports household-level data. Most questions are based on responses from 5 households per village (resulting in a sample of 2540 households). Some questions also include responses from an additional 15 households per village that were administered a short user feedback survey on recent health episodes, service provision, and satisfaction (resulting in an additional sample of 7620 households). The third panel reports individual-level data, from all individuals in a sampled household. Lower numbers of observations reflect missing data at baseline, the inability to resurvey a household at endline, or conditionality of variables on other responses. The last row shows p-values from Joint Orthogonality Tests. We aggregate all variables to the clinic (the highest common level) and estimate a Multinomial Logit where the dependent variable is the treatment group and the explanatory variables are all those in this table. In Columns (2)-(3) we set control as the base group and test the null that the coefficients on the explanatory variables are jointly zero for CM and NFA. In column (4) we re-estimate the model instead setting NFA as the base group, and test the null that the coefficients on CM are jointly zero. Significance: * is significant at the 10% level, ** is significant at the 5% level and *** is significant at the 1% level.

# D. Effects Prior to the Ebola Crisis

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

**Table D.1:** General Utilization

| | (1) Control Mean | (2) Pooled | (3) CM | (4) NFA | (5) Difference | (6) N |
|---|---|---|---|---|---|---|
| **General utilization** (SUR) | | 0.112 | 0.126 | 0.099 | 0.026 | 4496 |
| | | (0.031)*** | (0.034)*** | (0.036)*** | (0.033) | |
| | | [0.003]*** | [0.003]*** | [0.025]** | [0.748] | |
| (1) Number of health episodes in which sought Western care | 0.962 | 0.044 | 0.049 | 0.039 | 0.010 | 4496 |
| | (0.393) | (0.012)*** | (0.013)*** | (0.014)*** | (0.013) | |
| | | [0.001]*** | [0.001]*** | [0.008]*** | [0.765] | |

*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. The SUR treatment effects are estimated as an average of the z-scored treatment effects in this table using Seemingly Unrelated Regressions and controlling for average baseline observations of the outcomes at the community level, and including matching-triplet fixed effects. Column 1 reports standard deviation in parentheses. Columns 2–5 report robust standard errors, clustered by clinic. Multiple-inference corrected $q$-values that adjust for the false discovery rate within treatment arm across all outcomes in this table are shown in square brackets. Health episodes include questions about ante and post natal care, vaccinations and illness/injury episodes (asked for the past one month) and questions about child birth (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) Difference | (6) N |
|---|---|---|---|---|---|---|
| **Maternal utilization index** (SUR) | | 0.046 | 0.130 | −0.031 | 0.161 | 888 |
| | | (0.045) | (0.054)** | (0.054) | (0.057)*** | |
| | | [0.322] | [0.079]* | [0.463] | [0.044]** | |
| (1) ANC/PNC visits index | 0.000 | −0.038 | 0.008 | −0.079 | 0.087 | 887 |
| | (1.000) | (0.064) | (0.082) | (0.068) | (0.076) | |
| | | [0.386] | [0.861] | [0.968] | [0.147] | |
| (2) Birth in Western style facility | 0.834 | 0.048 | 0.094 | 0.006 | 0.088 | 877 |
| | (0.373) | (0.025)* | (0.028)*** | (0.030) | (0.031)*** | |
| | | [0.114] | [0.002]*** | [0.968] | [0.012]** | |

*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. The SUR treatment effects are estimated as an average of the z-scored treatment effects in this table using Seemingly Unrelated Regressions and controlling for average baseline observations of the outcomes at the community level, and including matching-triplet fixed effects. Column 1 reports standard deviation in parentheses. Columns 2–5 report robust standard errors, clustered by clinic. Multiple-inference corrected $q$-values that adjust for the false discovery rate within treatment arm across all outcomes in this table 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:** Satisfaction

| | (1) Control Mean | (2) Pooled | (3) CM | (4) NFA | (5) Difference | (6) N |
|---|---|---|---|---|---|---|
| **Satisfaction index** (SUR) | | 0.062 | 0.056 | 0.068 | −0.012 | 5052 |
| | | (0.029)** | (0.033)* | (0.033)** | (0.033) | |
| | | [0.059]* | [0.146] | [0.079]* | [0.863] | |
| (1) Satisfaction with family health | 3.439 | 0.053 | 0.047 | 0.058 | −0.011 | 5052 |
| | (0.670) | (0.026)** | (0.029) | (0.029)** | (0.027) | |
| | | [0.156] | [0.623] | [0.101] | [1.000] | |
| (2) Satisfaction with public health workers | 3.258 | 0.061 | 0.042 | 0.080 | −0.038 | 4994 |
| | (0.802) | (0.033)* | (0.040) | (0.039)** | (0.042) | |
| | | [0.156] | [0.623] | [0.101] | [1.000] | |
| (3) Satisfaction with care | 3.646 | 0.039 | 0.045 | 0.032 | 0.013 | 2535 |
| | (0.696) | (0.034) | (0.039) | (0.039) | (0.036) | |
| | | [0.193] | [0.623] | [0.258] | [1.000] | |
| (4) Individual would return to clinic | 0.967 | 0.007 | 0.006 | 0.007 | −0.001 | 2527 |
| | (0.174) | (0.007) | (0.008) | (0.008) | (0.008) | |
| | | [0.193] | [0.623] | [0.258] | [1.000] | |

*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. The SUR treatment effects are estimated as an average of the z-scored treatment effects in this table using Seemingly Unrelated Regressions and controlling for average baseline observations of the outcomes at the community level, and including matching-triplet fixed effects. Column 1 reports standard deviation in parentheses. Columns 2–5 report robust standard errors, clustered by clinic. Multiple-inference corrected $q$-values that adjust for the false discovery rate within treatment arm across all outcomes in this table 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.4:** Health Outcomes

| | (1) Control Mean | (2) Pooled | (3) CM | (4) NFA | (5) Difference | (6) N |
|---|---|---|---|---|---|---|
| **Health outcomes index** (SUR) | | 0.021 | 0.055 | −0.010 | 0.064 | 5053 |
| | | (0.023) | (0.025)** | (0.027) | (0.024)*** | |
| | | [0.322] | [0.094]* | [0.463] | [0.044]** | |
| (1) U5 Child death per HH | 0.039 | −0.009 | −0.015 | −0.004 | −0.011 | 5053 |
| | (0.193) | (0.005)* | (0.005)*** | (0.006) | (0.005)** | |
| | | [0.544] | [0.039]** | [1.000] | [0.071]* | |
| (2) Maternal death per HH | 0.001 | 0.000 | 0.001 | −0.001 | 0.001 | 5053 |
| | (0.035) | (0.001) | (0.001) | (0.001) | (0.001) | |
| | | [1.000] | [0.462] | [1.000] | [0.299] | |
| (3) Illness/injury in HH | 0.579 | −0.003 | −0.009 | 0.004 | −0.012 | 5053 |
| | (0.494) | (0.016) | (0.018) | (0.019) | (0.019) | |
| | | [1.000] | [0.462] | [1.000] | [0.672] | |
| (4) Child weight for length | 0.546 | 0.133 | 0.156 | 0.109 | 0.048 | 1991 |
| | (1.682) | (0.081) | (0.093)* | (0.093) | (0.093) | |
| | | [0.544] | [0.252] | [1.000] | [0.672] | |
| (5) Vaccine completion index (Under 2) | 3.085 | 0.032 | 0.303 | −0.209 | 0.512 | 1457 |
| | (2.560) | (0.152) | (0.184) | (0.163) | (0.164)*** | |
| | | [1.000] | [0.252] | [1.000] | [0.015]** | |
| (6) Child birth complication index | 0.000 | −0.026 | −0.126 | 0.061 | −0.187 | 856 |
| | (1.000) | (0.077) | (0.086) | (0.087) | (0.079)** | |
| | | [1.000] | [0.277] | [1.000] | [0.060]* | |
| (7) Child illness index | 0.002 | 0.024 | 0.031 | 0.018 | 0.012 | 4993 |
| | (0.995) | (0.102) | (0.115) | (0.115) | (0.107) | |
| | | [1.000] | [0.513] | [1.000] | [0.672] | |

*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. The SUR treatment effects are estimated as an average of the z-scored treatment effects in this table using Seemingly Unrelated Regressions and controlling for average baseline observations of the outcomes at the community level, and including matching-triplet fixed effects. Column 1 reports standard deviation in parentheses. Columns 2–5 report robust standard errors, clustered by clinic. Multiple-inference corrected $q$-values that adjust for the false discovery rate within treatment arm across all outcomes in this table are shown in square brackets. Maternal death is defined as death relating to either pregnancy complications or childbirth. Vaccine completion includes vaccines against 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.5:** Health Service Delivery

| | (1) Control Mean | (2) Pooled | (3) CM | (4) NFA | (5) Difference | (6) N |
|---|---|---|---|---|---|---|
| **Health service delivery index** (SUR) | | 0.014 | 0.022 | 0.012 | 0.010 | 2877 |
| | | (0.028) | (0.039) | (0.030) | (0.035) | |
| | | [0.465] | [0.476] | [0.463] | [0.863] | |
| (1) Absenteeism index | 0.000 | 0.098 | 0.030 | 0.120 | −0.090 | 2874 |
| | (1.000) | (0.055)* | (0.079) | (0.061)* | (0.084) | |
| | | [1.000] | [1.000] | [0.442] | [0.842] | |
| (2) Paid for treatment | 0.416 | −0.014 | −0.077 | 0.007 | −0.083 | 2872 |
| | (0.493) | (0.023) | (0.033)** | (0.023) | (0.030)*** | |
| | | [1.000] | [0.333] | [1.000] | [0.088]* | |
| (3) Amount paid | 8520.158 | 164.006 | −2812.243 | 1137.405 | −3949.648 | 2843 |
| | (31895.827) | (1484.834) | (2206.513) | (1475.317) | (1887.448)** | |
| | | [1.000] | [1.000] | [1.000] | [0.159] | |
| (4) Any problem | 0.063 | 0.003 | 0.009 | 0.001 | 0.008 | 2869 |
| | (0.242) | (0.015) | (0.020) | (0.016) | (0.017) | |
| | | [1.000] | [1.000] | [1.000] | [0.910] | |
| (5) Staff not present | 0.019 | 0.004 | −0.000 | 0.006 | −0.006 | 2869 |
| | (0.137) | (0.008) | (0.010) | (0.009) | (0.011) | |
| | | [1.000] | [1.000] | [1.000] | [0.910] | |
| (6) Drugs not available | 0.031 | −0.002 | 0.003 | −0.004 | 0.007 | 2869 |
| | (0.174) | (0.009) | (0.012) | (0.009) | (0.009) | |
| | | [1.000] | [1.000] | [1.000] | [0.910] | |
| (7) Facility not clean | 0.003 | −0.002 | −0.000 | −0.002 | 0.001 | 2869 |
| | (0.058) | (0.002) | (0.002) | (0.002) | (0.002) | |
| | | [1.000] | [1.000] | [1.000] | [0.910] | |
| (8) Unpleasant staff behavior | 0.022 | −0.006 | 0.006 | −0.010 | 0.016 | 2869 |
| | (0.148) | (0.007) | (0.011) | (0.007) | (0.009)* | |
| | | [1.000] | [1.000] | [1.000] | [0.209] | |
| (9) Medicine always in stock | 0.952 | 0.009 | 0.014 | 0.008 | 0.006 | 2478 |
| | (0.214) | (0.009) | (0.013) | (0.009) | (0.013) | |
| | | [1.000] | [1.000] | [1.000] | [0.910] | |
| (10) Individual satisfaction with care | 3.645 | 0.034 | 0.062 | 0.025 | 0.037 | 2863 |
| | (0.703) | (0.040) | (0.055) | (0.041) | (0.045) | |
| | | [1.000] | [1.000] | [1.000] | [0.910] | |
| (11) Individual would return to clinic | 0.969 | 0.008 | 0.006 | 0.008 | −0.003 | 2853 |
| | (0.173) | (0.007) | (0.011) | (0.007) | (0.010) | |
| | | [1.000] | [1.000] | [1.000] | [1.000] | |
| (12) Staff attitude | 3.735 | 0.042 | −0.030 | 0.065 | −0.095 | 2845 |
| | (0.580) | (0.029) | (0.045) | (0.030)** | (0.044)** | |
| | | [1.000] | [1.000] | [0.442] | [0.159] | |

*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. The SUR treatment effects are estimated as an average of the z-scored treatment effects in this table using Seemingly Unrelated Regressions and controlling for average baseline observations of the outcomes at the community level, and including matching-triplet fixed effects. Column 1 reports standard deviation in parentheses. Columns 2–5 report robust standard errors, clustered by clinic. Multiple-inference corrected $q$-values that adjust for the false discovery rate within treatment arm across all outcomes in this table 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.6:** Clinic Organization and Services

| | (1) Control Mean | (2) Pooled | (3) CM | (4) NFA | (5) Difference | (6) N |
|---|---|---|---|---|---|---|
| **Clinic organization and services index** (SUR) | | 0.055 | −0.002 | 0.112 | −0.114 | 254 |
| | | (0.064) | (0.074) | (0.075) | (0.078) | |
| | | [0.322] | [0.644] | [0.157] | [0.350] | |
| (1) Clinic service provision index | 0.000 | 0.149 | 0.095 | 0.203 | −0.108 | 254 |
| | (1.000) | (0.153) | (0.180) | (0.175) | (0.181) | |
| | | [1.000] | [1.000] | [1.000] | [1.000] | |
| (2) Clinic aware of free health care | 0.798 | 0.005 | −0.023 | 0.033 | −0.057 | 254 |
| | (0.404) | (0.052) | (0.060) | (0.058) | (0.056) | |
| | | [1.000] | [1.000] | [1.000] | [1.000] | |
| (3) Employee satisfaction index | 0.000 | 0.003 | −0.043 | 0.049 | −0.092 | 254 |
| | (1.000) | (0.127) | (0.141) | (0.152) | (0.146) | |
| | | [1.000] | [1.000] | [1.000] | [1.000] | |

*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. The SUR treatment effects are estimated as an average of the z-scored treatment effects in this table using Seemingly Unrelated Regressions and controlling for average baseline observations of the outcomes at the community level, and including matching-triplet fixed effects. Column 1 reports standard deviation in parentheses. Column 2-5 report robust standard errors. Multiple-inference corrected $q$-values that adjust for the false discovery rate within treatment arm across all outcomes in this table 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.7:** Community Support

| | (1) Control Mean | (2) Pooled | (3) CM | (4) NFA | (5) Difference | (6) N |
|---|---|---|---|---|---|---|
| **Contributions to clinic index** (SUR) | | 0.025 | 0.034 | 0.016 | 0.018 | 508 |
| | | (0.052) | (0.061) | (0.060) | (0.061) | |
| | | [0.465] | [0.476] | [0.463] | [0.863] | |
| (1) Contributions to clinic index (community survey) | 0.000 | 0.109 | 0.134 | 0.084 | 0.050 | 508 |
| | (1.000) | (0.078) | (0.092) | (0.092) | (0.096) | |
| | | [0.490] | [0.404] | [1.000] | [1.000] | |
| (2) Contributions to clinic index (facility survey) | 0.000 | −0.059 | −0.067 | −0.052 | −0.015 | 508 |
| | (0.997) | (0.121) | (0.142) | (0.136) | (0.138) | |
| | | [0.490] | [0.471] | [1.000] | [1.000] | |

*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. The SUR treatment effects are estimated as an average of the z-scored treatment effects in this table using Seemingly Unrelated Regressions and controlling for average baseline observations of the outcomes at the community level, and including matching-triplet fixed effects. Column 1 reports standard deviation in parentheses. Column 2-5 report robust standard errors, clustered by clinic. Multiple-inference corrected $q$-values that adjust for the false discovery rate within treatment arm across all outcomes in this table are shown in square brackets. 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) Difference | (6) N |
|---|---|---|---|---|---|---|
| **CDPE index** (SUR) | | 0.132 | 0.116 | 0.150 | −0.033 | 508 |
| | | (0.049)*** | (0.060)* | (0.056)*** | (0.062) | |
| | | [0.033]** | [0.114] | [0.025]** | [0.863] | |
| (1) Projects with local council/chief | 0.054 | 0.047 | 0.050 | 0.043 | 0.007 | 507 |
| | (0.226) | (0.022)** | (0.026)* | (0.026)* | (0.029) | |
| | | [0.090]* | [0.465] | [0.129] | [1.000] | |
| (2) Community provided labor | 0.042 | 0.040 | 0.038 | 0.042 | −0.004 | 504 |
| | (0.201) | (0.019)** | (0.022)* | (0.023)* | (0.025) | |
| | | [0.090]* | [0.465] | [0.129] | [1.000] | |
| (3) Community involved in planning | 0.030 | 0.022 | 0.026 | 0.018 | 0.008 | 504 |
| | (0.171) | (0.016) | (0.020) | (0.020) | (0.023) | |
| | | [0.103] | [0.465] | [0.161] | [1.000] | |
| (4) Problem addressed collectively? | 0.583 | −0.021 | −0.024 | −0.018 | −0.006 | 508 |
| | (0.494) | (0.040) | (0.045) | (0.046) | (0.046) | |
| | | [0.277] | [0.533] | [0.247] | [1.000] | |
| (5) Proportion has voter card | 0.983 | 0.005 | 0.003 | 0.007 | −0.004 | 489 |
| | (0.045) | (0.003) | (0.004) | (0.004)* | (0.004) | |
| | | [0.103] | [0.533] | [0.129] | [1.000] | |
| (6) Proportion voted in local election | 0.973 | 0.009 | 0.005 | 0.012 | −0.007 | 489 |
| | (0.054) | (0.004)** | (0.005) | (0.005)** | (0.005) | |
| | | [0.090]* | [0.465] | [0.099]* | [1.000] | |
| (7) Proportion voted in general election | 0.973 | 0.009 | 0.008 | 0.011 | −0.004 | 489 |
| | (0.055) | (0.005)** | (0.005) | (0.005)** | (0.005) | |
| | | [0.090]* | [0.465] | [0.109] | [1.000] | |

*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. The SUR treatment effects are estimated as an average of the z-scored treatment effects in this table using Seemingly Unrelated Regressions and controlling for average baseline observations of the outcomes at the community level, and including matching-triplet fixed effects. Column 1 reports standard deviation in parentheses. Columns 2–5 report robust standard errors, clustered by clinic. Multiple-inference corrected $q$-values that adjust for the false discovery rate within treatment arm across all outcomes in this table are shown in square brackets. 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) Difference | (6) N |
|---|---|---|---|---|---|---|
| **Water and sanitation index** (SUR) | | 0.110 | 0.066 | 0.154 | −0.088 | 5053 |
| | | (0.044)** | (0.051) | (0.051)*** | (0.050)* | |
| | | [0.037]** | [0.224] | [0.025]** | [0.261] | |
| (1) Water and sanitation HH index | 0.000 | 0.068 | −0.022 | 0.159 | −0.180 | 5053 |
| | (1.000) | (0.058) | (0.068) | (0.063)** | (0.062)*** | |
| | | [0.089]* | [0.362] | [0.040]** | [0.013]** | |
| (2) Water and sanitation village index | −0.016 | 0.175 | 0.132 | 0.218 | −0.087 | 5053 |
| | (1.003) | (0.087)** | (0.097) | (0.104)** | (0.102) | |
| | | [0.073]* | [0.298] | [0.040]** | [0.658] | |
| (3) Satisfaction with village sanitation index | 0.000 | 0.087 | 0.087 | 0.086 | 0.001 | 5051 |
| | (1.000) | (0.043)** | (0.049)* | (0.050)* | (0.051) | |
| | | [0.073]* | [0.298] | [0.059]* | [1.000] | |

*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. The SUR treatment effects are estimated as an average of the z-scored treatment effects in this table using Seemingly Unrelated Regressions and controlling for average baseline observations of the outcomes at the community level, and including matching-triplet fixed effects. Column 1 reports standard deviation in parentheses. Columns 2–5 report robust standard errors, clustered by clinic. Multiple-inference corrected $q$-values that adjust for the false discovery rate within treatment arm across all outcomes in this table are shown in square brackets. 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) Difference | (6) N |
|---|---|---|---|---|---|---|
| **Economic outcomes index** (SUR) | | 0.053 | 0.036 | 0.070 | −0.034 | 5053 |
| | | (0.032) | (0.037) | (0.040)* | (0.042) | |
| | | [0.138] | [0.311] | [0.116] | [0.748] | |
| (1) Physical asset index | 0.000 | 0.020 | −0.023 | 0.063 | −0.086 | 5052 |
| | (1.000) | (0.048) | (0.053) | (0.059) | (0.060) | |
| | | [1.000] | [1.000] | [0.404] | [0.249] | |
| (2) Agricultural asset index | 0.000 | 0.120 | 0.174 | 0.065 | 0.109 | 5051 |
| | (1.000) | (0.044)*** | (0.057)*** | (0.050) | (0.063)* | |
| | | [0.026]** | [0.010]*** | [0.404] | [0.249] | |
| (3) Dwelling materials index | 0.000 | 0.053 | 0.020 | 0.087 | −0.067 | 5052 |
| | (1.000) | (0.053) | (0.060) | (0.058) | (0.055) | |
| | | [0.875] | [1.000] | [0.404] | [0.249] | |
| (4) Total consumption expenditure | 0.000 | 0.019 | −0.027 | 0.066 | −0.093 | 5053 |
| | (1.000) | (0.045) | (0.050) | (0.055) | (0.056)* | |
| | | [1.000] | [1.000] | [0.404] | [0.249] | |

*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. The SUR treatment effects are estimated as an average of the z-scored treatment effects in this table using Seemingly Unrelated Regressions and controlling for average baseline observations of the outcomes at the community level, and including matching-triplet fixed effects. Column 1 reports standard deviation in parentheses. Columns 2–5 report robust standard errors, clustered by clinic. Multiple-inference corrected $q$-values that adjust for the false discovery rate within treatment arm across all outcomes in this table 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.

## 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) Difference | (6) N |
|---|---|---|---|---|---|---|
| **Maternal utilization index** (SUR) | | 0.046 | 0.130 | −0.031 | 0.161 | 888 |
| | | (0.045) | (0.054)** | (0.054) | (0.057)*** | |
| | | [0.322] | [0.079]* | [0.463] | [0.044]** | |
| (1) ANC/PNC visits index | 0.000 | −0.038 | 0.008 | −0.079 | 0.087 | 887 |
| | (1.000) | (0.064) | (0.082) | (0.068) | (0.076) | |
| | | [0.386] | [0.861] | [0.968] | [0.147] | |
| (2) Birth in Western medicine facility | 0.000 | 0.130 | 0.252 | 0.017 | 0.236 | 877 |
| | (1.000) | (0.066)* | (0.074)*** | (0.082) | (0.084)*** | |
| | | [0.114] | [0.002]*** | [0.968] | [0.012]** | |

*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. The SUR treatment effects are estimated as an average of the z-scored treatment effects in this table using Seemingly Unrelated Regressions and controlling for average baseline observations of the outcomes at the community level, and including matching-triplet fixed effects. Column 1 reports standard deviation in parentheses. Columns 2–5 report robust standard errors, clustered by clinic. Multiple-inference corrected $q$-values that adjust for the false discovery rate within treatment arm across all outcomes in this table are shown in square brackets. Variables are control-group normalized at endline (z-scored). Significance: * is significant at the 10% level, ** is significant at the 5% level and *** is significant at the 1% level.

**Table D.12:** Satisfaction (z-scored)

| | (1) Control Mean | (2) Pooled | (3) CM | (4) NFA | (5) Difference | (6) N |
|---|---|---|---|---|---|---|
| **Satisfaction index** (SUR) | | 0.062 | 0.056 | 0.068 | −0.012 | 5052 |
| | | (0.029)** | (0.033)* | (0.033)** | (0.033) | |
| | | [0.058]* | [0.139] | [0.080]* | [0.863] | |
| (1) Satisfaction with family health | 0.000 | 0.079 | 0.071 | 0.087 | −0.016 | 5052 |
| | (1.000) | (0.039)** | (0.044) | (0.043)** | (0.040) | |
| | | [0.156] | [0.623] | [0.101] | [1.000] | |
| (2) Satisfaction with public health workers | 0.000 | 0.076 | 0.053 | 0.100 | −0.047 | 4994 |
| | (1.000) | (0.042)* | (0.050) | (0.049)** | (0.052) | |
| | | [0.156] | [0.623] | [0.101] | [1.000] | |
| (3) Satisfaction with care | 0.000 | 0.056 | 0.065 | 0.046 | 0.019 | 2535 |
| | (1.000) | (0.049) | (0.055) | (0.056) | (0.052) | |
| | | [0.193] | [0.623] | [0.258] | [1.000] | |
| (4) Individual would return to clinic | 0.000 | 0.038 | 0.036 | 0.040 | −0.004 | 2527 |
| | (1.000) | (0.039) | (0.046) | (0.043) | (0.046) | |
| | | [0.193] | [0.623] | [0.258] | [1.000] | |

*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. The SUR treatment effects are estimated as an average of the z-scored treatment effects in this table using Seemingly Unrelated Regressions and controlling for average baseline observations of the outcomes at the community level, and including matching-triplet fixed effects. Column 1 reports standard deviation in parentheses. Columns 2–5 report robust standard errors, clustered by clinic. Multiple-inference corrected $q$-values that adjust for the false discovery rate within treatment arm across all outcomes in this table are shown in square brackets. Variables are control-group normalized at endline (z-scored). Significance: * is significant at the 10% level, ** is significant at the 5% level and *** is significant at the 1% level.

**Table D.13:** Health outcomes (z-scored)

| | (1) Control Mean | (2) Pooled | (3) CM | (4) NFA | (5) Difference | (6) N |
|---|---|---|---|---|---|---|
| **Health outcomes index** (SUR) | | 0.021 | 0.055 | −0.010 | 0.064 | 5053 |
| | | (0.023) | (0.025)** | (0.027) | (0.024)*** | |
| | | [0.322] | [0.094]* | [0.463] | [0.044]** | |
| (1) U5 Child death per HH | 0.000 | −0.047 | −0.075 | −0.019 | −0.056 | 5053 |
| | (1.000) | (0.025)* | (0.027)*** | (0.030) | (0.027)** | |
| | | [0.544] | [0.039]** | [1.000] | [0.071]* | |
| (2) Maternal death per HH | 0.000 | 0.000 | 0.017 | −0.017 | 0.034 | 5053 |
| | (1.000) | (0.026) | (0.031) | (0.028) | (0.028) | |
| | | [1.000] | [0.462] | [1.000] | [0.299] | |
| (3) Illness/injury in HH | 0.000 | −0.005 | −0.018 | 0.007 | −0.025 | 5053 |
| | (1.000) | (0.033) | (0.037) | (0.039) | (0.038) | |
| | | [1.000] | [0.462] | [1.000] | [0.672] | |
| (4) Child weight for length | 0.000 | 0.079 | 0.093 | 0.065 | 0.028 | 1991 |
| | (1.000) | (0.048) | (0.055)* | (0.055) | (0.055) | |
| | | [0.544] | [0.252] | [1.000] | [0.672] | |
| (5) Vaccine completion index (Under 2) | 0.000 | 0.012 | 0.118 | −0.082 | 0.200 | 1457 |
| | (1.000) | (0.059) | (0.072) | (0.064) | (0.064)*** | |
| | | [1.000] | [0.252] | [1.000] | [0.015]** | |
| (6) Child birth complication index | 0.000 | −0.026 | −0.126 | 0.061 | −0.187 | 856 |
| | (1.000) | (0.077) | (0.086) | (0.087) | (0.079)** | |
| | | [1.000] | [0.277] | [1.000] | [0.060]* | |
| (7) Child illness index | 0.000 | 0.024 | 0.031 | 0.018 | 0.012 | 4993 |
| | (1.000) | (0.102) | (0.115) | (0.116) | (0.108) | |
| | | [1.000] | [0.513] | [1.000] | [0.672] | |

*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. The SUR treatment effects are estimated as an average of the z-scored treatment effects in this table using Seemingly Unrelated Regressions and controlling for average baseline observations of the outcomes at the community level, and including matching-triplet fixed effects. Column 1 reports standard deviation in parentheses. Columns 2–5 report robust standard errors, clustered by clinic. Multiple-inference corrected $q$-values that adjust for the false discovery rate within treatment arm across all outcomes in this table are shown in square brackets. Maternal death is defined as death relating to either pregnancy complications or childbirth. Vaccine completion includes vaccines against diphtheria, tetanus, whooping cough, hepatitis B, and haemophilus influenza type B. Variables are control-group normalized at endline (z-scored). Significance: * is significant at the 10% level, ** is significant at the 5% level and *** is significant at the 1% level.

**Table D.14:** Health Service Delivery (z-scored)

| | (1) Control Mean | (2) Pooled | (3) CM | (4) NFA | (5) Difference | (6) N |
|---|---|---|---|---|---|---|
| **Health service delivery index** (SUR) | | 0.014 | 0.022 | 0.012 | 0.010 | 2877 |
| | | (0.028) | (0.039) | (0.030) | (0.035) | |
| | | [0.355] | [0.254] | [0.463] | [0.589] | |
| (1) Absenteeism index | 0.000 | 0.098 | 0.030 | 0.120 | −0.090 | 2874 |
| | (1.000) | (0.055)* | (0.079) | (0.061)* | (0.084) | |
| | | [1.000] | [1.000] | [0.442] | [0.842] | |
| (2) Paid for treatment | 0.000 | −0.028 | −0.155 | 0.014 | −0.169 | 2872 |
| | (1.000) | (0.047) | (0.067)** | (0.047) | (0.062)*** | |
| | | [1.000] | [0.333] | [1.000] | [0.088]* | |
| (3) Amount paid | 0.000 | 0.005 | −0.088 | 0.036 | −0.124 | 2843 |
| | (1.000) | (0.047) | (0.069) | (0.046) | (0.059)** | |
| | | [1.000] | [1.000] | [1.000] | [0.159] | |
| (4) Any problem | 0.000 | 0.011 | 0.035 | 0.003 | 0.032 | 2869 |
| | (1.000) | (0.064) | (0.083) | (0.066) | (0.072) | |
| | | [1.000] | [1.000] | [1.000] | [0.910] | |
| (5) Staff not present | 0.000 | 0.031 | −0.002 | 0.043 | −0.045 | 2869 |
| | (1.000) | (0.057) | (0.076) | (0.062) | (0.078) | |
| | | [1.000] | [1.000] | [1.000] | [0.910] | |
| (6) Drugs not available | 0.000 | −0.011 | 0.019 | −0.021 | 0.040 | 2869 |
| | (1.000) | (0.053) | (0.068) | (0.054) | (0.052) | |
| | | [1.000] | [1.000] | [1.000] | [0.910] | |
| (7) Facility not clean | 0.000 | −0.027 | −0.008 | −0.033 | 0.025 | 2869 |
| | (1.000) | (0.027) | (0.036) | (0.029) | (0.037) | |
| | | [1.000] | [1.000] | [1.000] | [0.910] | |
| (8) Unpleasant staff behavior | 0.000 | −0.043 | 0.040 | −0.071 | 0.110 | 2869 |
| | (1.000) | (0.051) | (0.074) | (0.050) | (0.062)* | |
| | | [1.000] | [1.000] | [1.000] | [0.209] | |
| (9) Medicine always in stock | 0.000 | 0.042 | 0.064 | 0.036 | 0.028 | 2478 |
| | (1.000) | (0.041) | (0.062) | (0.044) | (0.062) | |
| | | [1.000] | [1.000] | [1.000] | [0.910] | |
| (10) Individual satisfaction with care | 0.000 | 0.048 | 0.088 | 0.035 | 0.052 | 2863 |
| | (1.000) | (0.057) | (0.078) | (0.058) | (0.065) | |
| | | [1.000] | [1.000] | [1.000] | [0.910] | |
| (11) Individual would return to clinic | 0.000 | 0.044 | 0.032 | 0.048 | −0.016 | 2853 |
| | (1.000) | (0.042) | (0.066) | (0.041) | (0.059) | |
| | | [1.000] | [1.000] | [1.000] | [1.000] | |
| (12) Staff attitude | 0.000 | 0.072 | −0.051 | 0.112 | −0.163 | 2845 |
| | (1.000) | (0.050) | (0.077) | (0.052)** | (0.075)** | |
| | | [1.000] | [1.000] | [0.442] | [0.159] | |

*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. The SUR treatment effects are estimated as an average of the z-scored treatment effects in this table using Seemingly Unrelated Regressions and controlling for average baseline observations of the outcomes at the community level, and including matching-triplet fixed effects. Column 1 reports standard deviation in parentheses. Columns 2–5 report robust standard errors, clustered by clinic. Multiple-inference corrected $q$-values that adjust for the false discovery rate within treatment arm across all outcomes in this table 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. Variables are control-group normalized at endline (z-scored). Significance: * is significant at the 10% level, ** is significant at the 5% level and *** is significant at the 1% level.

**Table D.15:** Clinic Organization and Services (z-scored)

| | (1) Control Mean | (2) Pooled | (3) CM | (4) NFA | (5) Difference | (6) N |
|---|---|---|---|---|---|---|
| **Clinic organization and services index** (SUR) | | 0.055 | −0.002 | 0.112 | −0.114 | 254 |
| | | (0.064) | (0.074) | (0.075) | (0.078) | |
| | | [0.322] | [0.644] | [0.157] | [0.350] | |
| (1) Clinic service provision index | 0.000 | 0.149 | 0.095 | 0.203 | −0.108 | 254 |
| | (1.000) | (0.153) | (0.180) | (0.175) | (0.181) | |
| | | [1.000] | [1.000] | [1.000] | [1.000] | |
| (2) Clinic aware of free health care | 0.000 | 0.012 | −0.058 | 0.083 | −0.140 | 254 |
| | (1.000) | (0.129) | (0.148) | (0.144) | (0.138) | |
| | | [1.000] | [1.000] | [1.000] | [1.000] | |
| (3) Employee satisfaction index | 0.000 | 0.003 | −0.043 | 0.049 | −0.092 | 254 |
| | (1.000) | (0.127) | (0.141) | (0.152) | (0.146) | |
| | | [1.000] | [1.000] | [1.000] | [1.000] | |

*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. The SUR treatment effects are estimated as an average of the z-scored treatment effects in this table using Seemingly Unrelated Regressions and controlling for average baseline observations of the outcomes at the community level, and including matching-triplet fixed effects. Column 1 reports standard deviation in parentheses. Columns 2–5 report robust standard errors. Multiple-inference corrected $q$-values that adjust for the false discovery rate within treatment arm across all outcomes in this table 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. Variables are control-group normalized at endline (z-scored). Significance: * is significant at the 10% level, ** is significant at the 5% level and *** is significant at the 1% level.

**Table D.16:** Community Support (z-scored)

| | (1) Control Mean | (2) Pooled | (3) CM | (4) NFA | (5) Difference | (6) N |
|---|---|---|---|---|---|---|
| **Contributions to clinic index** (SUR) | | 0.025 | 0.034 | 0.016 | 0.018 | 508 |
| | | (0.052) | (0.061) | (0.060) | (0.061) | |
| | | [0.465] | [0.476] | [0.463] | [0.863] | |
| (1) Contributions to clinic index (community survey) | 0.000 | 0.109 | 0.134 | 0.084 | 0.050 | 508 |
| | (1.000) | (0.078) | (0.092) | (0.092) | (0.096) | |
| | | [0.490] | [0.404] | [1.000] | [1.000] | |
| (2) Contributions to clinic index (facility survey) | 0.000 | −0.060 | −0.067 | −0.052 | −0.015 | 508 |
| | (1.000) | (0.121) | (0.143) | (0.136) | (0.139) | |
| | | [0.490] | [0.471] | [1.000] | [1.000] | |

*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. The SUR treatment effects are estimated as an average of the z-scored treatment effects in this table using Seemingly Unrelated Regressions and controlling for average baseline observations of the outcomes at the community level, and including matching-triplet fixed effects. Column 1 reports standard deviation in parentheses. Columns 2–5 report robust standard errors, clustered by clinic. Multiple-inference corrected $q$-values that adjust for the false discovery rate within treatment arm across all outcomes in this table are shown in square brackets. 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. Variables are control-group normalized at endline (z-scored). 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) Difference | (6) N |
|---|---|---|---|---|---|---|
| **CDPE index** (SUR) | | 0.132 | 0.116 | 0.150 | −0.033 | 508 |
| | | (0.049)*** | (0.060)* | (0.056)*** | (0.062) | |
| | | [0.033]** | [0.114] | [0.025]** | [0.790] | |
| (1) Projects with local council/chief | 0.000 | 0.208 | 0.223 | 0.192 | 0.032 | 507 |
| | (1.000) | (0.095)** | (0.114)* | (0.116)* | (0.127) | |
| | | [0.090]* | [0.465] | [0.129] | [1.000] | |
| (2) Community provided labor | 0.000 | 0.200 | 0.190 | 0.209 | −0.019 | 504 |
| | (1.000) | (0.095)** | (0.112)* | (0.116)* | (0.126) | |
| | | [0.090]* | [0.465] | [0.129] | [1.000] | |
| (3) Community involved in planning | 0.000 | 0.127 | 0.150 | 0.103 | 0.047 | 504 |
| | (1.000) | (0.095) | (0.117) | (0.115) | (0.132) | |
| | | [0.103] | [0.465] | [0.161] | [1.000] | |
| (4) Problem addressed collectively? | 0.000 | −0.043 | −0.049 | −0.037 | −0.012 | 508 |
| | (1.000) | (0.080) | (0.092) | (0.094) | (0.093) | |
| | | [0.277] | [0.533] | [0.247] | [1.000] | |
| (5) Proportion has voter card | 0.000 | 0.102 | 0.061 | 0.145 | −0.084 | 489 |
| | (1.000) | (0.076) | (0.087) | (0.087)* | (0.084) | |
| | | [0.103] | [0.533] | [0.129] | [1.000] | |
| (6) Proportion voted in local election | 0.000 | 0.163 | 0.099 | 0.231 | −0.131 | 489 |
| | (1.000) | (0.082)** | (0.094) | (0.092)** | (0.090) | |
| | | [0.090]* | [0.465] | [0.099]* | [1.000] | |
| (7) Proportion voted in general election | 0.000 | 0.172 | 0.140 | 0.206 | −0.066 | 489 |
| | (1.000) | (0.085)** | (0.097) | (0.096)** | (0.091) | |
| | | [0.090]* | [0.465] | [0.109] | [1.000] | |

*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. The SUR treatment effects are estimated as an average of the z-scored treatment effects in this table using Seemingly Unrelated Regressions and controlling for average baseline observations of the outcomes at the community level, and including matching-triplet fixed effects. Column 1 reports standard deviation in parentheses. Columns 2–5 report robust standard errors, clustered by clinic. Multiple-inference corrected $q$-values that adjust for the false discovery rate within treatment arm across all outcomes in this table are shown in square brackets. 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. Variables are control-group normalized at endline (z-scored). 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) Difference | (6) N |
|---|---|---|---|---|---|---|
| **Water and sanitation index** (SUR) | | 0.110 | 0.066 | 0.154 | −0.088 | 5053 |
| | | (0.044)** | (0.051) | (0.051)*** | (0.050)* | |
| | | [0.037]** | [0.210] | [0.025]** | [0.261] | |
| (1) Water and sanitation HH index | 0.000 | 0.068 | −0.022 | 0.159 | −0.180 | 5053 |
| | (1.000) | (0.058) | (0.068) | (0.063)** | (0.062)*** | |
| | | [0.089]* | [0.362] | [0.040]** | [0.013]** | |
| (2) Water and sanitation village index | 0.000 | 0.174 | 0.131 | 0.218 | −0.086 | 5053 |
| | (1.000) | (0.086)** | (0.097) | (0.104)** | (0.102) | |
| | | [0.073]* | [0.298] | [0.040]** | [0.658] | |
| (3) Satisfaction with village sanitation index | 0.000 | 0.087 | 0.087 | 0.086 | 0.001 | 5051 |
| | (1.000) | (0.043)** | (0.049)* | (0.050)* | (0.051) | |
| | | [0.073]* | [0.298] | [0.059]* | [1.000] | |

*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. The SUR treatment effects are estimated as an average of the z-scored treatment effects in this table using Seemingly Unrelated Regressions and controlling for average baseline observations of the outcomes at the community level, and including matching-triplet fixed effects. Column 1 reports standard deviation in parentheses. Columns 2–5 report robust standard errors, clustered by clinic. Multiple-inference corrected $q$-values that adjust for the false discovery rate within treatment arm across all outcomes in this table are shown in square brackets. 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. Variables are control-group normalized at endline (z-scored). 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) Difference | (6) N |
|---|---|---|---|---|---|---|
| **Economic outcomes index** (SUR) | | 0.053 | 0.036 | 0.070 | −0.034 | 5053 |
| | | (0.032) | (0.037) | (0.040)* | (0.042) | |
| | | [0.138] | [0.262] | [0.119] | [0.589] | |
| (1) Physical asset index | 0.000 | 0.020 | −0.023 | 0.063 | −0.086 | 5052 |
| | (1.000) | (0.048) | (0.053) | (0.059) | (0.060) | |
| | | [1.000] | [1.000] | [0.404] | [0.249] | |
| (2) Agricultural asset index | 0.000 | 0.120 | 0.174 | 0.065 | 0.109 | 5051 |
| | (1.000) | (0.044)*** | (0.057)*** | (0.050) | (0.063)* | |
| | | [0.026]** | [0.010]*** | [0.404] | [0.249] | |
| (3) Dwelling materials index | 0.000 | 0.053 | 0.020 | 0.087 | −0.067 | 5052 |
| | (1.000) | (0.053) | (0.060) | (0.058) | (0.055) | |
| | | [0.875] | [1.000] | [0.404] | [0.249] | |
| (4) Total consumption expenditure | 0.000 | 0.019 | −0.027 | 0.066 | −0.093 | 5053 |
| | (1.000) | (0.045) | (0.050) | (0.055) | (0.056)* | |
| | | [1.000] | [1.000] | [0.404] | [0.249] | |

*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. The SUR treatment effects are estimated as an average of the z-scored treatment effects in this table using Seemingly Unrelated Regressions and controlling for average baseline observations of the outcomes at the community level, and including matching-triplet fixed effects. Column 1 reports standard deviation in parentheses. Column 2-5 report robust standard errors, clustered by clinic. Multiple-inference corrected $q$-values that adjust for the false discovery rate within treatment arm across all outcomes in this table are shown in square brackets. Variables are control-group normalized at endline (z-scored). 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:** PHU utilization

| | (1) Control Mean | (2) Pooled | (3) CM | (4) NFA | (5) Difference | (6) N |
|---|---|---|---|---|---|---|
| (1) Number of health episodes in which sought Western care | 0.962 | 0.044 | 0.049 | 0.039 | 0.010 | 4496 |
| | (0.393) | (0.012)*** | (0.013)*** | (0.014)*** | (0.013) | |
| (2) Number of health episodes in which sought Western care at PHU | 0.883 | 0.058 | 0.064 | 0.052 | 0.012 | 4496 |
| | (0.463) | (0.017)*** | (0.019)*** | (0.020)** | (0.018) | |

*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. Column 1 reports standard deviation in parentheses. Columns 2–5 report robust standard errors, clustered by clinic. Significance: * is significant at the 10% level, ** is significant at the 5% level and *** is significant at the 1% level.

**Table D.21:** Utilization of Western and traditional medicine among illness and injury patients

| | (1) Control Mean | (2) Pooled | (3) CM | (4) NFA | (5) Difference | (6) N |
|---|---|---|---|---|---|---|
| (1) Number of illness/injury episodes in which sought Western care | 0.899 | 0.026 | 0.029 | 0.023 | 0.006 | 2617 |
| | (0.301) | (0.013)** | (0.015)* | (0.015) | (0.014) | |
| (2) Number of illness/injury episodes in which sought traditional or religious care | 0.073 | −0.019 | −0.015 | −0.024 | 0.009 | 2612 |
| | (0.260) | (0.011)* | (0.012) | (0.012)** | (0.011) | |

*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. Column 1 reports standard deviation in parentheses. Columns 2–5 report robust standard errors, clustered by clinic. Significance: * is significant at the 10% level, ** is significant at the 5% level and *** is significant at the 1% level.

**Table D.22:** Utilization of Western and traditional medicine among illness and injury patients (z-scored)

| | (1) Control Mean | (2) Pooled | (3) CM | (4) NFA | (5) Difference | (6) N |
|---|---|---|---|---|---|---|
| (1) Number of illness/injury episodes in which sought Western care | 0.000 | 0.087 | 0.097 | 0.076 | 0.021 | 2617 |
| | (1.000) | (0.044)** | (0.050)* | (0.049) | (0.046) | |
| (2) Number of illness/injury episodes in which sought traditional or religious care | 0.000 | −0.074 | −0.057 | −0.092 | 0.036 | 2612 |
| | (1.000) | (0.043)* | (0.048) | (0.047)** | (0.042) | |

*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.23:** 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. Column 1 reports standard deviation in parentheses. Columns 2–4 report robust standard errors, clustered by clinic, in parentheses. Significance: * is significant at the 10% level, ** is significant at the 5% level and *** is significant at the 1% level.

**Table D.24:** Main outcome families when controlling for baseline imbalances

| | (1) Control Mean | (2) Pooled | (3) CM | (4) NFA | (5) Difference | (6) N |
|---|---|---|---|---|---|---|
| General utilization | 0.000 | 0.095 | 0.105 | 0.083 | 0.022 | 4451 |
| | (1.000) | (0.033)*** | (0.035)*** | (0.040)** | (0.035) | |
| | | [0.048]** | [0.015]** | [0.144] | [1.000] | |
| Maternal utilization index | 0.000 | 0.064 | 0.178 | −0.053 | 0.232 | 878 |
| | (1.000) | (0.070) | (0.081)** | (0.083) | (0.087)*** | |
| | | [0.349] | [0.083]* | [0.596] | [0.038]** | |
| Health outcomes index | 0.000 | 0.084 | 0.171 | −0.008 | 0.179 | 5003 |
| | (1.000) | (0.051)* | (0.055)*** | (0.059) | (0.056)*** | |
| | | [0.182] | [0.015]** | [0.956] | [0.016]** | |
| Satisfaction index | 0.000 | 0.103 | 0.088 | 0.119 | −0.030 | 5002 |
| | (1.000) | (0.042)** | (0.049)* | (0.048)** | (0.048) | |
| | | [0.078]* | [0.147] | [0.144] | [1.000] | |
| Health service delivery index | 0.000 | 0.048 | 0.073 | 0.040 | 0.033 | 2845 |
| | (1.000) | (0.059) | (0.082) | (0.062) | (0.075) | |
| | | [0.349] | [0.479] | [0.596] | [1.000] | |
| Clinic organization and services index | 0.000 | 0.149 | 0.056 | 0.257 | −0.201 | 254 |
| | (1.000) | (0.158) | (0.182) | (0.187) | (0.192) | |
| | | [0.349] | [0.612] | [0.313] | [1.000] | |
| CDPE index | 0.000 | 0.156 | 0.147 | 0.165 | −0.019 | 501 |
| | (1.000) | (0.087)* | (0.104) | (0.103) | (0.113) | |
| | | [0.182] | [0.241] | [0.238] | [1.000] | |
| Contributions to clinic index | 0.000 | 0.027 | 0.049 | 0.003 | 0.046 | 501 |
| | (1.000) | (0.097) | (0.114) | (0.111) | (0.115) | |
| | | [0.641] | [0.587] | [0.956] | [1.000] | |
| Water and sanitation index | 0.000 | 0.108 | 0.061 | 0.158 | −0.097 | 5003 |
| | (1.000) | (0.062)* | (0.071) | (0.072)** | (0.071) | |
| | | [0.182] | [0.479] | [0.144] | [0.820] | |
| Economic outcomes index | 0.000 | 0.037 | 0.046 | 0.028 | 0.018 | 5003 |
| | (1.000) | (0.052) | (0.060) | (0.062) | (0.064) | |
| | | [0.357] | [0.479] | [0.686] | [1.000] | |

*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. Column 1 reports standard deviation in parentheses. Columns 2–5 report robust standard errors, clustered by clinic. Multiple-inference corrected $q$-values that adjust for the false discovery rate within treatment arm across all outcomes in this table are shown in square brackets. In this table, we control for baseline variables displaying imbalance in Appendix Table C.3 – 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, education, 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 the1% level.

**Table D.25:** Effects on record keeping

| | (1) Control Mean | (2) Pooled | (3) CM | (4) NFA | (5) Difference | (6) N |
|---|---|---|---|---|---|---|
| (1) Clinic registers available | 0.988 | −0.012 | −0.024 | −0.000 | −0.023 | 254 |
| | (0.109) | (0.018) | (0.022) | (0.019) | (0.018) | |
| (2) Tally sheets available | 0.929 | 0.006 | −0.006 | 0.018 | −0.024 | 254 |
| | (0.259) | (0.034) | (0.040) | (0.039) | (0.038) | |
| (3) DHIS monthly report available | 0.988 | −0.000 | −0.012 | 0.012 | −0.024 | 254 |
| | (0.109) | (0.015) | (0.019) | (0.015) | (0.017) | |
| (4) Medicine stock cards available | 0.952 | −0.012 | −0.036 | 0.012 | −0.048 | 254 |
| | (0.214) | (0.032) | (0.038) | (0.035) | (0.037) | |
| (5) Cumulative coverage record for ante-natal care is available | 0.452 | 0.011 | −0.025 | 0.047 | −0.073 | 254 |
| | (0.501) | (0.066) | (0.076) | (0.076) | (0.074) | |
| (6) Cumulative coverage record for BCG vaccine is available | 0.417 | 0.016 | −0.027 | 0.059 | −0.086 | 254 |
| | (0.496) | (0.065) | (0.074) | (0.076) | (0.073) | |
| (7) Cumulative coverage record for measles vaccine is available | 0.417 | 0.023 | −0.026 | 0.071 | −0.098 | 254 |
| | (0.496) | (0.065) | (0.073) | (0.075) | (0.072) | |
| (8) Cumulative coverage record for penta3 vaccine is available | 0.393 | 0.059 | −0.002 | 0.119 | −0.121 | 254 |
| | (0.491) | (0.066) | (0.075) | (0.075) | (0.073) | |

*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. Column 1 reports standard deviation in parentheses. Columns 2–5 report robust standard errors. Significance: * is significant at the 10% level, ** is significant at the 5% level and *** is significant at the 1% level.

**Table D.26:** Perceived quality of care in full sample

| | (1) Control Mean | (2) Pooled | (3) CM | (4) NFA | (5) Difference | (6) N |
|---|---|---|---|---|---|---|
| (1) **Perceived quality of care index** | 0.000 | 0.150 | 0.149 | 0.151 | −0.003 | 5041 |
| | (1.000) | (0.039)*** | (0.045)*** | (0.044)*** | (0.045) | |
| (2) General utilization | 0.945 | 0.043 | 0.046 | 0.041 | 0.005 | 3330 |
| | (0.349) | (0.012)*** | (0.014)*** | (0.014)*** | (0.013) | |
| (3) Satisfaction with public health workers | 3.258 | 0.056 | 0.036 | 0.076 | −0.039 | 4994 |
| | (0.802) | (0.034)* | (0.040) | (0.039)* | (0.042) | |
| (4) Effectiveness of Western medicine relative to traditional or religious healing | −0.350 | 0.038 | 0.046 | 0.029 | 0.017 | 4290 |
| | (0.578) | (0.022)* | (0.025)* | (0.025) | (0.024) | |

*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. Column 1 reports standard deviation in parentheses. Columns 2–5 report robust standard errors, clustered by clinic. Significance: * is significant at the 10% level, ** is significant at the 5% level and *** is significant at the 1% level.

# E. Longer-run Effects during the Ebola Crisis

## E.1 Training for Health Care Workers on Infection Prevention and Control

**Table E.1:** 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 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.[A8] 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.[A9]

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.[A10] 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 Appendix Table E.2).

[A8] 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.

[A9] http://wiki.openstreetmap.org/wiki/2014_West_Africa_Ebola_Response

[A10] We use optimal string alignment distance, a variant of the Levenshtein distance, which is commonly employed in geo-coding algorithms.

**Table E.2:** 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.3 Descriptive Statistics for Reported Ebola Cases

**Table E.3:** Descriptive Statistics for Ebola Sample

| Measure | Control | Pooled Treatment |
|---|---|---|
| **Total Ebola Cases** | | |
| Section-week observations | 1836 | 3604 |
| Sum | 515 | 1530 |
| Average | 0.281 | 0.425 |
| Std. Dev. | 0.727 | 1.282 |
| Minimum | 0 | 0 |
| Maximum | 6 | 31 |
| Proportion with no cases | 0.818 | 0.789 |
| Proportion with cases | 0.182 | 0.211 |
| **Confirmed Ebola Cases** | | |
| Section-week observations | 1836 | 3604 |
| Sum | 21 | 248 |
| Average | 0.011 | 0.069 |
| Std. Dev. | 0.129 | 0.687 |
| Minimum | 0 | 0 |
| Maximum | 3 | 28 |
| Proportion with no cases | 0.991 | 0.97 |
| Proportion with cases | 0.009 | 0.03 |

*Notes*: Descriptive statistics for total and confirmed reported Ebola cases in control and pooled (CM or NFA) treatment arms. The sample includes 160 sections over 34 weeks.

## E.4 Baseline Balance in Ebola Sample

**Table E.4:** Baseline Balance (Ebola sub-sample)

| | (1) Control Mean | (2) CM–Control Difference | (3) NFA–Control Difference | (4) CM–NFA Difference | (5) N |
|---|---|---|---|---|---|
| *Village characteristics* | | | | | |
| Motorable road | 0.863 | 0.108 | 0.045 | 0.063 | 318 |
| | (0.345) | (0.059)* | (0.050) | (0.055) | |
| Mobile phone coverage within 1 mile from the community | 0.764 | 0.078 | 0.174 | −0.096 | 318 |
| | (0.426) | (0.076) | (0.058)*** | (0.065) | |
| Distance to the closest clinic | 1.390 | 0.338 | 0.909 | −0.571 | 318 |
| | (2.757) | (0.532) | (0.905) | (0.505) | |
| Travel cost to closest clinic | 164.835 | −29.666 | −46.167 | 16.501 | 317 |
| | (1163.570) | (132.127) | (124.628) | (102.778) | |
| *Household characteristics and questions to household head* | | | | | |
| Household size | 4.445 | −0.023 | 0.051 | −0.074 | 3021 |
| | (3.086) | (0.075) | (0.073) | (0.083) | |
| Number of illness or injury cases per household | 0.069 | −0.039 | −0.016 | −0.024 | 3021 |
| | (0.267) | (0.017)** | (0.016) | (0.018) | |
| Birth in household last year | 0.166 | −0.012 | 0.016 | −0.028 | 1340 |
| | (0.372) | (0.020) | (0.021) | (0.020) | |
| Child under 2 in household | 0.246 | −0.022 | 0.004 | −0.027 | 1339 |
| | (0.431) | (0.028) | (0.025) | (0.026) | |
| Prominent village member in household | 0.046 | −0.028 | −0.030 | 0.002 | 1331 |
| | (0.210) | (0.014)** | (0.013)** | (0.014) | |
| Believes doctor's advice | 0.993 | 0.003 | −0.003 | 0.006 | 1235 |
| | (0.083) | (0.006) | (0.007) | (0.007) | |
| Health care fees unaffordable | 2.352 | −0.028 | 0.031 | −0.059 | 1289 |
| | (0.814) | (0.077) | (0.078) | (0.082) | |
| Trust in the community | 1.902 | −0.111 | −0.038 | −0.072 | 1339 |
| | (0.679) | (0.065)* | (0.057) | (0.073) | |
| Community cohesion | 2.375 | −0.020 | 0.035 | −0.055 | 1336 |
| | (0.595) | (0.047) | (0.050) | (0.054) | |
| Believe VHC members represent your interest | 2.932 | −0.115 | −0.186 | 0.072 | 618 |
| | (0.992) | (0.127) | (0.132) | (0.114) | |
| The VHC can be trusted | 2.373 | −0.098 | 0.131 | −0.229 | 725 |
| | (0.895) | (0.121) | (0.109) | (0.119)* | |
| *Individual characteristics* | | | | | |
| Muslim | 0.835 | 0.003 | 0.007 | −0.005 | 6168 |
| | (0.371) | (0.042) | (0.037) | (0.047) | |
| Mende (Ethnicity) | 0.508 | −0.036 | 0.006 | −0.042 | 6167 |
| | (0.500) | (0.017)** | (0.016) | (0.018)** | |
| Temne (Ethnicity) | 0.266 | 0.098 | 0.140 | −0.042 | 6167 |
| | (0.442) | (0.057)* | (0.046)*** | (0.057) | |
| Highest level of education | 1.667 | 0.146 | 0.657 | −0.511 | 5924 |
| | (2.838) | (0.151) | (0.146)*** | (0.160)*** | |
| *Chi*-square *p*-value | | 0.656 | 0.233 | 0.626 | |

*Notes:* This table presents baseline balance for households included in both baseline and endline surveys in the Ebola subsample, the 160 sections that contain a single clinic, and hence unique treatment status. Column (1) shows the mean and standard deviation of the control group at baseline. Columns (2) and (3) show regression coefficients on the CM and NFA treatment arm indicators compared to control, with standard errors clustered at the clinic level in parentheses. Column (4) compares the CM treatment arm to the NFA arm. Column (5) displays the sample size. The second panel reports data at the household level. Most questions are based on a sample of 5 households per village; others also include an additional 15 households that were administered a short user feedback survey on recent health episodes, service provision, and satisfaction. The third panel reports data at the individual level, from all individuals in a sampled household. Lower numbers of observations relative to the number of sampled households reflect missing data at baseline, the inability to resurvey a household at endline, or conditionality of variables on other responses. The last row shows p-values from Joint Orthogonality Tests. We aggregate all variables to the clinic (the highest common level) and estimate a Multinomial Logit where the dependent variable is the treatment group and the explanatory variables are all those in this table. In Columns (2)-(3) we set control as the base group and test the null that the coefficients on the explanatory variables are jointly zero for CM and NFA. In column (4) we re-estimate the model instead setting NFA as the base group, and test the null that the coefficients on CM are jointly zero. Significance: * is significant at the 10% level, ** is significant at the 5% level and *** is significant at the 1% level.

## E.5 Extending Panel to August 2014

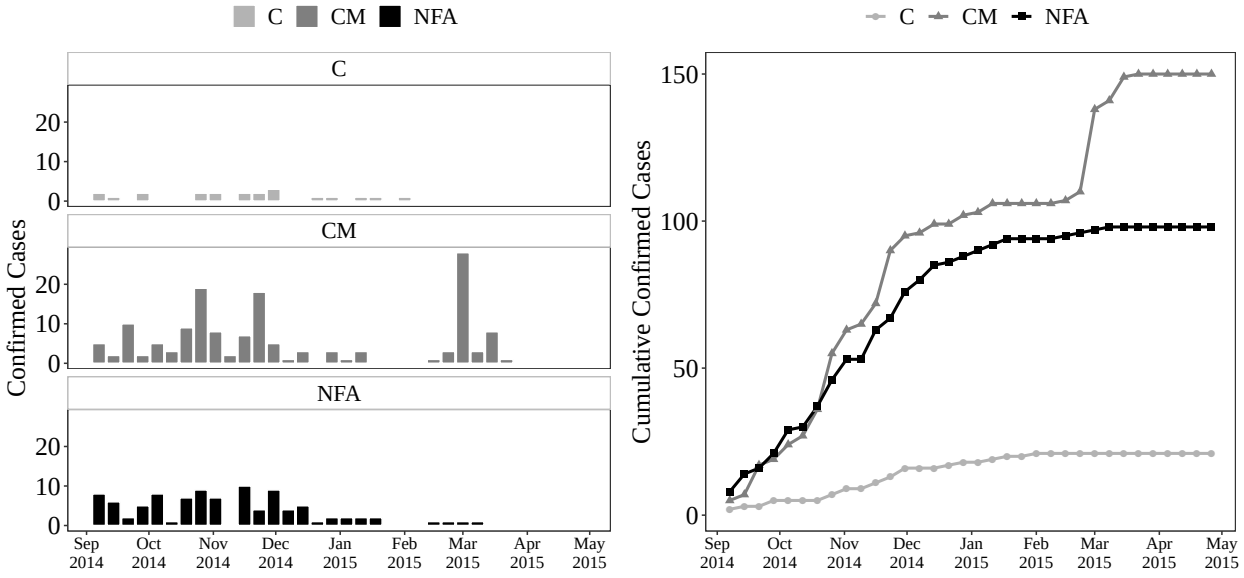
**Table E.5:** Effect on Reported Cases in Extended Panel (August 2014–April 2015)

| | 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*: Treatment effects estimated using OLS including matching-triplet and week fixed effects. Column 1 reports standard deviation in parentheses. Column 2-4 report robust standard errors, clustered by section, in parentheses. Difference column reports the difference between the CM and NFA coefficients; the standard error is computed using the delta method. 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. Significance: * is significant at the 10% level, ** is significant at the 5% level and *** is significant at the 1% level.

## E.6 Time-series of Confirmed Cases

**Figure E.1:** Confirmed Ebola Cases by Treatment



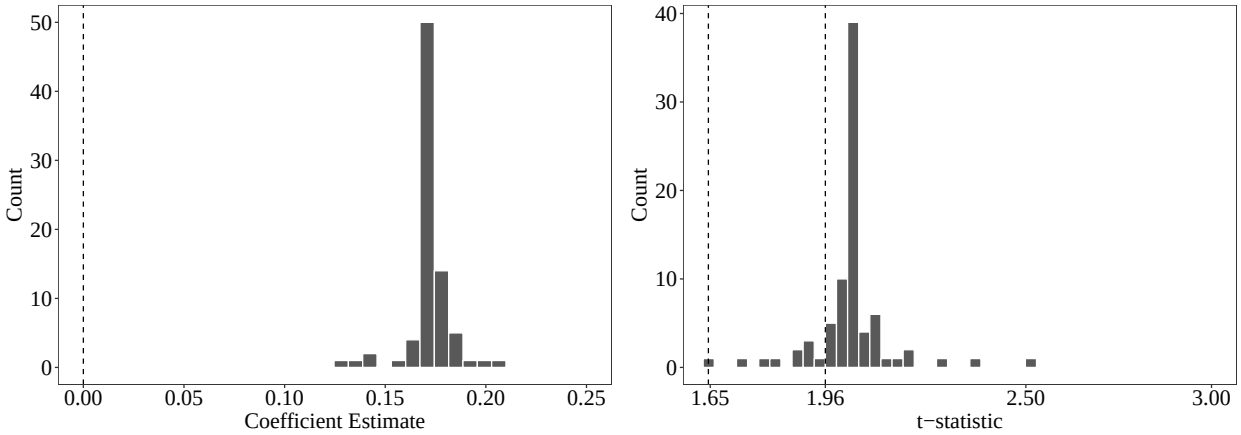

**(a)** Weekly Counts

**(b)** Cumulative Counts

Appendix Figure E.1(a) plots the time series of confirmed Ebola cases by week; bars represent the raw counts. C refers to control; CM refers to community monitoring; NFA refers to non-financial awards. We use the date that the case was first saved in the VHF, which is available for 96% of cases in our sample. Appendix Figure E.1(b) graphs the cumulative count of confirmed Ebola cases by treatment group.

## E.7 Dropping Triplets

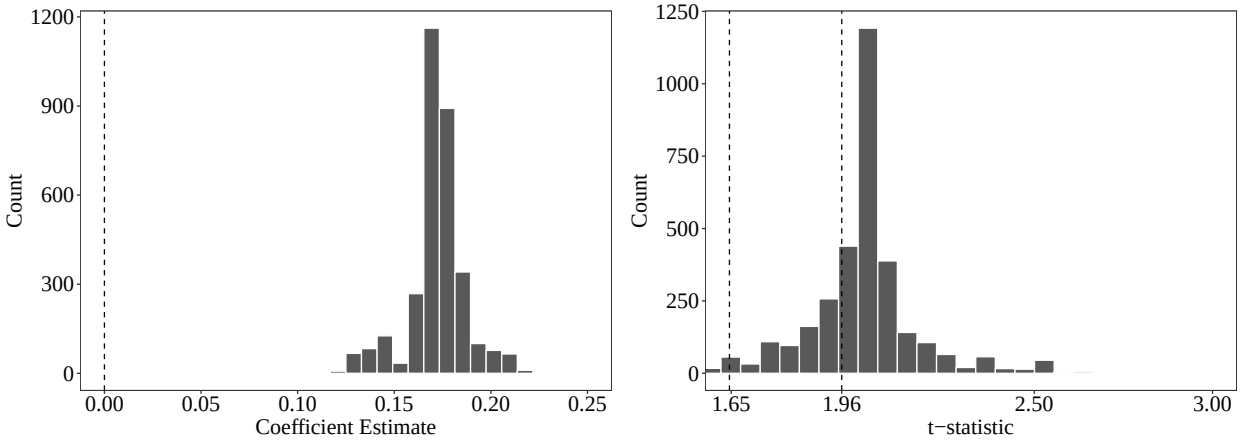
**Figure E.2:** Estimates Dropping Each Triplet from Sample




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

We re-estimate Equation 3 dropping one triplet (i.e., block) from the sample with each iteration. Appendix Figure E.2(a): distribution of coefficient estimates. Appendix Figure E.2(b): distribution of t-statistics.

**Figure E.3:** Estimates Dropping Each Pair of Triplets from Sample

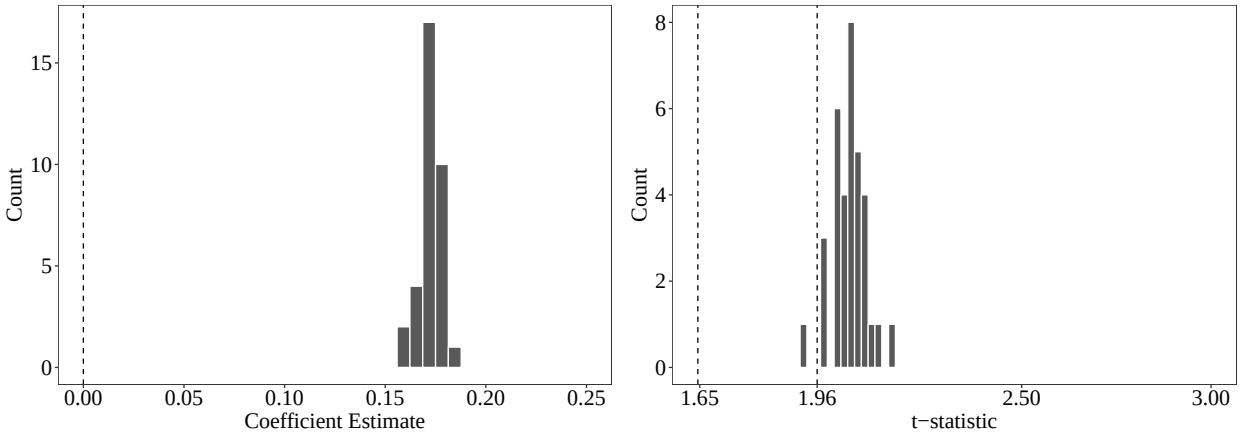



**(a)** Coefficient ($\widehat{\beta}$) **(b)** t-Statistic

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

## E.8 Dropping Weeks

**Figure E.4:** 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. Appendix Figure E.4(a): distribution of coefficient estimates. Appendix Figure E.4(b): distribution of t-statistics.

## E.9 Effect on Patient Deaths

**Table E.6:** Patient Deaths per Section per Week

| | Predicted Deaths | | | | | | |
|---|---|---|---|---|---|---|---|
| Total Cases in Last 2 Weeks | Control (C) (1) | Pooled (P) (2) | CM (3) | NFA (4) | C - P (5) | C - CM (6) | C - NFA (7) |
| 2 Cases | 0.49 | 0.36 | 0.31 | 0.45 | -0.13 | -0.18 | -0.05 |
| | (0.04) | (0.05) | (0.06) | (0.04) | (0.06)** | (0.07)** | (0.05) |
| 5 Cases | 1.23 | 0.8 | 0.6 | 1.13 | -0.43 | -0.63 | -0.1 |
| | (0.11) | (0.17) | (0.18) | (0.11) | (0.19)** | (0.2)*** | (0.14) |
| 10 Cases | 2.45 | 1.53 | 1.09 | 2.27 | -0.92 | -1.38 | -0.2 |
| | (0.21) | (0.36) | (0.39) | (0.23) | (0.4)** | (0.43)*** | (0.3) |

*Notes*: Treatment effects estimated using OLS including matching-triplet and week fixed effects. Robust standard errors, clustered by section, in parentheses Predicted deaths based on estimates in Table 4 model 1. Significance: * is significant at the 10% level, ** is significant at the 5% level and *** is significant at the 1% level.

**Table E.7:** 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. Column 1 reports standard deviation in parentheses. Column 2-4 report robust standard errors in parentheses. Difference column reports the difference between the CM and NFA coefficients; the standard error is computed using the delta method. Delays greater than 60 days were removed to limit the influence of outliers. Significance: * is significant at the 10% level, ** is significant at the 5% level and *** is significant at the 1% level.

## E.10 Calculating Reduction in the Reproduction Rate

$R_0$ 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.[A11] $R_0$ 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 an infected individual and susceptible individual coming into contact, the likelihood that Ebola is transmitted between the two is unaffected by treatment. Under 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 Appendix Table E.7).

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.[A12] Going forward, assume $r(T=0) = 0.5$ and $r(T=1) = r(T=0) \cdot \tau$, where $\tau$ is the treatment effect.

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 $\widehat{\tau} = (0.281 + 0.173)/(0.281) = 1.62$.

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 Community Care Centers contributed to a 13–32 percent reduction in $R_0$.[A13]

[A11] https://ajph.aphapublications.org/doi/suppl/10.2105/AJPH.2015.303020/suppl_file/web+appendix+r2.docx

[A12] https://stacks.cdc.gov/view/cdc/24901

[A13] Their estimate is likely conservative, as they do not incorporate how Community Care Centers affect reporting rates.

## E.11 Effect on Reported Cases by Month

**Table E.8:** Effect on Total Cases

| | Control Mean | Pooled | CM | NFA | Difference | N |
|---|---|---|---|---|---|---|
| **2014** | | | | | | |
| 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) | |
| **2015** | | | | | | |
| 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*: Treatment effects estimated using OLS including matching-triplet and week fixed effects. Column 1 reports standard deviation in parentheses. Column 2-4 report robust standard errors, clustered by section, in parentheses. Difference column reports the difference between the CM and NFA coefficients; the standard error is computed using the delta method. N varies because case counts are recorded weekly, and there can be 4 or 5 weeks within each period. Significance: * is significant at the 10% level, ** is significant at the 5% level and *** is significant at the 1% level.

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

For the 160 sections in our sample, we identify their nearest out-of-sample neighbor (using the distance between the centroids of sections). We then assign that nearest neighbor the treatment status of the in-sample section and re-run our analysis.

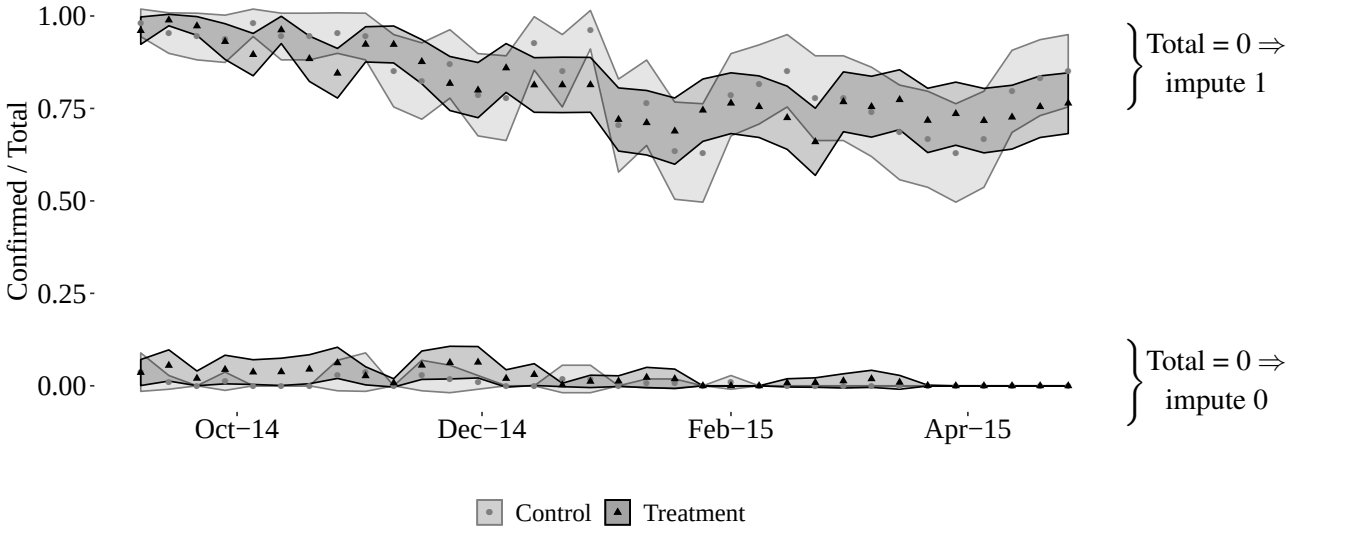
**Table E.9:** Placebo: Reported Cases using Nearest Neighboring Section

| | 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) | |

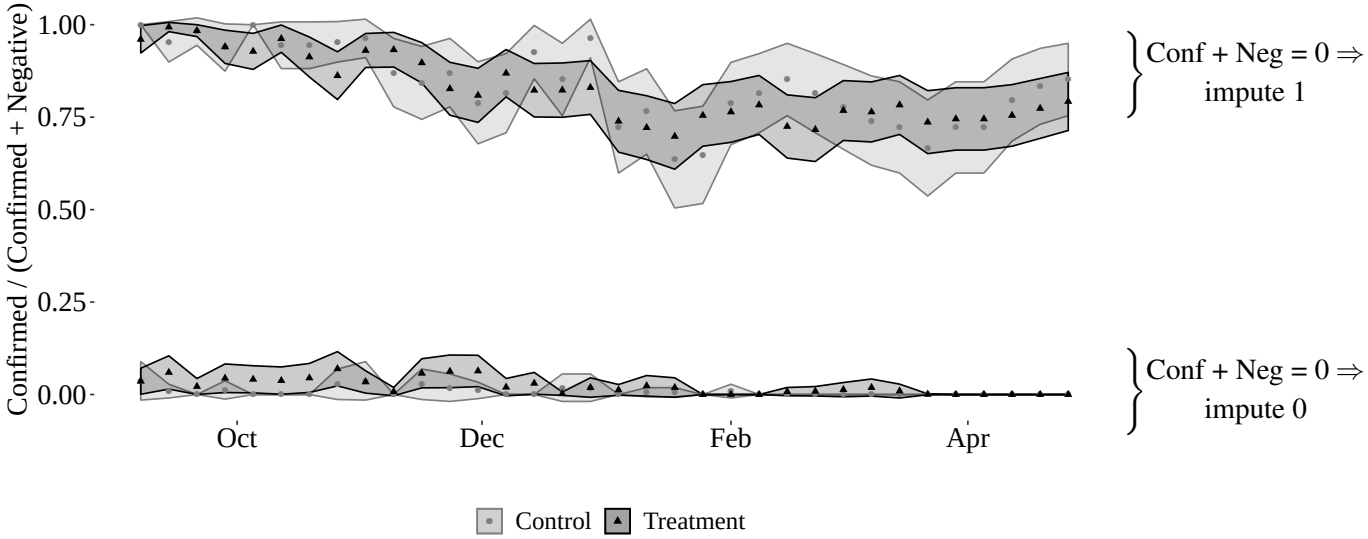
*Notes*: Treatment effects estimated using OLS including matching-triplet and week fixed effects. Column 1 reports standard deviation in parentheses. Column 2-4 report robust standard errors, clustered by section, in parentheses. Difference column reports the difference between the CM and NFA coefficients; the standard error is computed using the delta method. The three major cities in our study districts (Bo Town, Kenema Town, and Makeni Town) are excluded as potential nearest neighbors. Significance: * is significant at the 10% level, ** is significant at the 5% level and *** is significant at the 1% level.

## E.13 Ratio of Confirmed and Total Cases

**Figure E.5:** Ratio of Confirmed to Total Cases




**(a)** Ratio Confirmed to Total




**(b)** Ratio Confirmed to Confirmed and Negative

Appendix Figure E.5(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. Appendix Figure E.5(b) computes the ratio of confirmed to confirmed plus negative cases. If 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.14 Bounding Exercise: Unintended Increase

Data on Ebola incidence in Sierra Leone is incomplete. As a result, 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, $s \in \{0,1\}$; prior to testing, $i$ does not know with certainty if they are infected, $I \in \{0,1\}$. 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, $T \in \{0,1\}$. $i$ 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.[A14] The infection rates within control and treatment communities, $e_1 = \mathbb{E}(I \mid T = 1)$ and $e_0 = \mathbb{E}(I \mid T = 0)$, are also common knowledge.

$i$ must 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 $\tau_h \in \mathbb{R}^1_+$ is the treatment effect on reporting among symptomatic individuals
- 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 $\tau_l \in \mathbb{R}^1_+$ is the treatment effect on reporting among asymptomatic individuals

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.

### Logic

We estimate the percentage difference in confirmed cases between treatment and control $\beta$ where:

[A14] This is likely a considerable proportion of Sierra Leone's population: a 2016 assessment for example found that 27 percent of children under 5 had malaria in the *two weeks* prior to the survey (Leone et al. (2016)). Over months, the probability of flu-like symptoms due to illnesses unrelated to Ebola is quite likely.

$e_1$ $\overline{e_0\tau_h=\mathscr{E}\tau_h=\beta}$

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 $\tau_h$ 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 $\beta$ 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 (E.14) and substituting,

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

This implies

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

*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 ($\tau_l$). We set $\beta = \widehat{\beta} \approx 1.4$ and $e_1 = 0.01$.[A15] 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.
- This differential effect ($\tau_l / \tau_h$) 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 $\tau_l / \tau_h$ must be large.

[A15] 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 $\beta \tau_h / e_1$.

**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 $\tau_l$ must reach 1.67.

**Table E.10:** Implied $\mathscr{E}$ and $\tau_l$

| $p$ | $d$ | $\tau_h$ | $\tau_l$ | $\mathscr{E}$ |
|---|---|---|---|---|
| 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.[A16] (This would correspond to $h = 0.4$ in our terms.) This implies that $\tau_h$ 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 individuals were failing to seek care and, thus, might have changed their decision as a consequence of treatment.

In sum, 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

[A16] https://www.cdc.gov/mmwr/pdf/other/su6303.pdf

difficult to believe. Of course, this conclusion rests on taking seriously the patterns we observe in the data, in particular the constant ratio of confirmed to total cases regardless of treatment status.

## E.15 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."[A17]

**Table E.11:** Contact Tracing among Confirmed Patients by Treatment

| | | | Proportion among Contacts | |
|---|---|---|---|---|
| Treatment | Total Traced | Proportion Traced | 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.12:** 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. Column 1 reports standard deviation in parentheses. Column 2-4 report robust standard errors in parentheses. Difference column reports the difference between the CM and NFA coefficients; the standard error is computed using the delta method. Significance: * is significant at the 10% level, ** is significant at the 5% level and *** is significant at the 1% level.

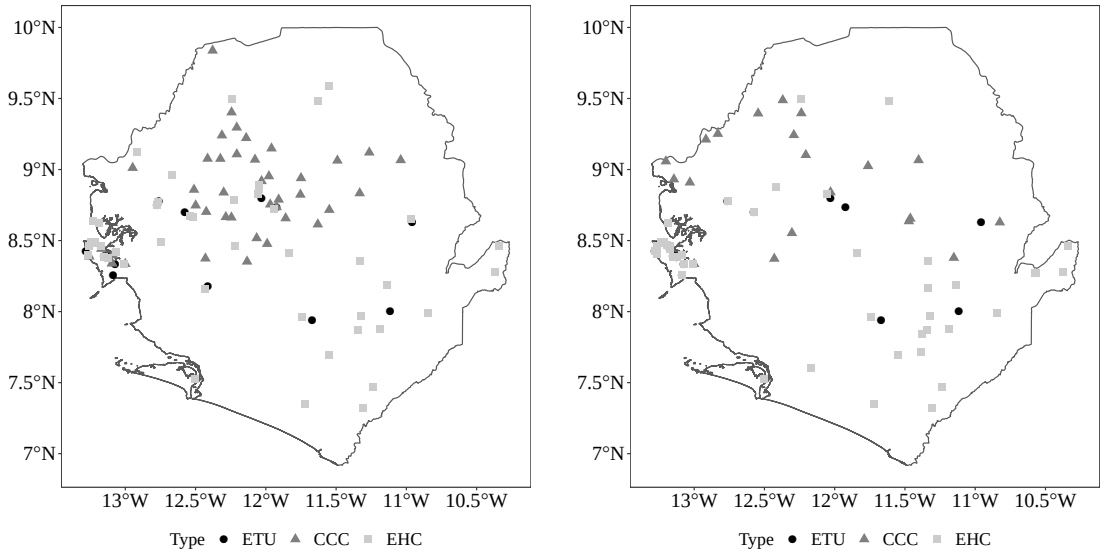
[A17] https://www.who.int/csr/resources/publications/ebola/contact-tracing-guidelines/en/

## E.16 Ebola-specific Balance Tests

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.6:** 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.13:** Balance: Specialized Ebola Facilities and Non-sample Clinics

| | Control Mean | Pooled | CM | NFA | Difference | N |
|---|---|---|---|---|---|---|
| NERC | | | | | | |
| Total | 0.056 | -0.03 | -0.054 | -0.011 | -0.043 | 160 |
| | (0.231) | (0.042) | (0.051) | (0.048) | (0.051) | |
| EHC | 0.019 | 0.006 | -0.016 | 0.023 | -0.04 | 160 |
| | (0.136) | (0.033) | (0.04) | (0.037) | (0.04) | |
| CCC | 0.019 | -0.018 | -0.027 | -0.011 | -0.015 | 160 |
| | (0.136) | (0.021) | (0.026) | (0.024) | (0.026) | |
| Beds | 0.463 | -0.18 | -0.66 | 0.194 | -0.854 | 160 |
| | (2.044) | (0.493) | (0.596) | (0.556) | (0.601) | |
| UNMEER | | | | | | |
| Total | 0.093 | 0.03 | 0.024 | 0.035 | -0.011 | 160 |
| | (0.293) | (0.065) | (0.079) | (0.074) | (0.08) | |
| EHC | 0.019 | -0.012 | -0.013 | -0.011 | -0.001 | 160 |
| | (0.136) | (0.029) | (0.035) | (0.033) | (0.036) | |
| CCC | 0.056 | 0.06 | 0.048 | 0.069 | -0.021 | 160 |
| | (0.231) | (0.056) | (0.068) | (0.064) | (0.069) | |
| Beds | 0.759 | 0.347 | 0.351 | 0.345 | 0.006 | 160 |
| | (2.495) | (0.623) | (0.762) | (0.712) | (0.769) | |
| MoHS | | | | | | |
| Non-Sample Clinics | 0.204 | -0.126 | -0.17 | -0.091 | -0.079 | 160 |
| | (0.762) | (0.093) | (0.114) | (0.106) | (0.115) | |

*Notes*: Differences estimated using OLS including matching-triplet fixed effects. Top panel uses data from the Sierra Leone National Ebola Response Center (NERC); middle panel from the UN Mission for Ebola Emergency Response (UNMEER); and the bottom panel from the Ministry of Health and Sanitation (MoHS). ETU refers to Ebola Treatment Unit; EHC, to Ebola Holding Center; CCC, to Community Care Center; and Non-Sample Clinics, to clinics not included in the RCT. Significance: * is significant at the 10% level, ** is significant at the 5% level and *** is significant at the 1% level.

**Table E.14:** Balance: Minimum Distance to Specialized Ebola Facilities

| | Control Mean | Pooled | CM | NFA | Difference | N |
|---|---|---|---|---|---|---|
| NERC | | | | | | |
| ETU | 33.986 | 0.588 | -1.302 | 2.056 | -3.358 | 160 |
| | (17.422) | (3.795) | (4.626) | (4.319) | (4.668) | |
| EHC | 20.715 | -2.034 | -0.315 | -3.369 | 3.054 | 160 |
| | (11.479) | (2.115) | (2.563) | (2.394) | (2.587) | |
| CCC | 49.794 | -5.353 | -3.257 | -6.98 | 3.723 | 160 |
| | (31.244) | (4.09) | (4.984) | (4.654) | (5.03) | |
| UNMEER | | | | | | |
| ETU | 33.57 | 0.82 | -1.156 | 2.354 | -3.511 | 160 |
| | (17.495) | (3.777) | (4.602) | (4.297) | (4.644) | |
| EHC | 20.586 | -2.361 | -1.105 | -3.337 | 2.232 | 160 |
| | (10.445) | (2.011) | (2.446) | (2.284) | (2.468) | |
| CCC | 54.163 | -7.277 | -5.159 | -8.921 | 3.762 | 160 |
| | (43.64) | (3.819)* | (4.651) | (4.343)** | (4.693) | |

*Notes*: Differences estimated using OLS including matching-triplet fixed effects. Top panel uses data from the Sierra Leone's National Ebola Response Center (NERC); bottom panel from the UN Mission for Ebola Emergency Response (UNMEER); ETU refers to Ebola Treatment Unit; EHC refers to Ebola Holding Center; CCC refers to Community Care Center. Significance: * is significant at the 10% level, ** is significant at the 5% level and *** is significant at the 1% level.

**Table E.15:** Balance: Proxies for Exposure

| | Control Mean | Pooled | CM | NFA | Difference | N |
|---|---|---|---|---|---|---|
| Dist(Patient Zero in Guinea) | 196.354 | 10.93 | 8.158 | 13.083 | -4.925 | 160 |
| | (42.021) | (5.079)** | (6.186) | (5.777)** | (6.242) | |
| Dist(Patient Zero in SL) | 91.104 | 11.627 | 7.758 | 14.632 | -6.873 | 160 |
| | (62.208) | (6.441)* | (7.838) | (7.319)** | (7.91) | |
| Primary Road Density | 0.007 | 0.008 | -0.012 | 0.023 | -0.035 | 160 |
| | (0.029) | (0.015) | (0.018) | (0.017) | (0.018)* | |
| Secondary Road Density | 0.019 | 0.011 | 0.013 | 0.01 | 0.002 | 160 |
| | (0.053) | (0.011) | (0.014) | (0.013) | (0.014) | |
| Tertiary Road Density | 0.084 | -0.009 | 0.01 | -0.024 | 0.034 | 160 |
| | (0.158) | (0.024) | (0.029) | (0.027) | (0.03) | |
| Ruggedness | 35.915 | -4.394 | -3.679 | -4.95 | 1.271 | 160 |
| | (34.693) | (6.405) | (7.832) | (7.313) | (7.903) | |
| Number of Rivers | 2.333 | 0.407 | 0.17 | 0.591 | -0.421 | 160 |
| | (1.197) | (0.344) | (0.418) | (0.39) | (0.422) | |

*Notes*: Differences estimated using OLS including matching-triplet fixed effects. Outcomes are at section level. Distance are in kilometers. Road density is road length normalized by area. Ruggedness is the standard deviation of elevation in meters. Significance: * is significant at the 10% level, ** is significant at the 5% level and *** is significant at the 1% level.

## E.17 Alternative Functional Forms for Reported Cases

**Table E.16:** 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) | |
| Rare Events Logit | | | | | | |
| Total | 0.182 | 0.332 | 0.417 | 0.273 | 0.144 | 5,440 |
| | (0.386) | (0.184)* | (0.245)* | (0.213) | (0.272) | |
| Confirmed | 0.009 | 1.114 | 1.156 | 1.086 | 0.07 | 5,440 |
| | (0.096) | (0.275)*** | (0.345)*** | (0.281)*** | (0.286) | |
| Negative | 0.164 | 0.293 | 0.307 | 0.283 | 0.024 | 5,440 |
| | (0.371) | (0.186) | (0.241) | (0.218) | (0.267) | |

*Notes*: Treatment effects estimated including matching-triplet and week fixed effects. Dependent variable in top and bottom panel is a dummy, indicating if a section-week reported any case. The top two panels are estimated using OLS, the third panel uses Poisson and final panel uses Rare Event Logit. Column 1 reports standard deviation in parentheses. Column 2-4 report robust standard errors, clustered by section, in parentheses. Difference column reports the difference between the CM and NFA coefficients; the standard error is computed using the delta method. Significance: * is significant at the 10% level, ** is significant at the 5% level and *** is significant at the 1% level.

## E.18 Cross-sectional Results for Reported Cases

**Table E.17:** 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) | (4.05) | (5.636) | (4.787) | (6.448) | |
| Confirmed | 0.389 | 2.012 | 2.909 | 1.315 | 1.594 | 160 |
| | (1.235) | (1.152)* | (1.817) | (1.23) | (1.97) | |
| Negative | 8.093 | 3.383 | 2.689 | 3.922 | -1.233 | 160 |
| | (10.617) | (2.909) | (3.728) | (3.64) | (4.481) | |
| Log(Ebola Cases + 1) | | | | | | |
| Total | 1.798 | 0.294 | 0.417 | 0.198 | 0.219 | 160 |
| | (1.11) | (0.22) | (0.291) | (0.255) | (0.317) | |
| Confirmed | 0.171 | 0.38 | 0.413 | 0.354 | 0.059 | 160 |
| | (0.456) | (0.142)*** | (0.204)** | (0.156)** | (0.214) | |
| Negative | 1.674 | 0.227 | 0.348 | 0.133 | 0.215 | 160 |
| | (1.07) | (0.212) | (0.266) | (0.249) | (0.292) | |
| Linear Probability Model | | | | | | |
| Total | 0.833 | 0.072 | 0.091 | 0.057 | 0.034 | 160 |
| | (0.376) | (0.068) | (0.091) | (0.073) | (0.088) | |
| Confirmed | 0.148 | 0.251 | 0.234 | 0.265 | -0.03 | 160 |
| | (0.359) | (0.083)*** | (0.114)** | (0.096)*** | (0.125) | |
| Negative | 0.833 | 0.048 | 0.096 | 0.011 | 0.086 | 160 |
| | (0.376) | (0.073) | (0.094) | (0.082) | (0.096) | |
| IHS(Ebola Cases) | | | | | | |
| Total | 2.242 | 0.335 | 0.484 | 0.22 | 0.264 | 160 |
| | (1.328) | (0.259) | (0.341) | (0.298) | (0.369) | |
| Confirmed | 0.221 | 0.474 | 0.505 | 0.45 | 0.055 | 160 |
| | (0.587) | (0.175)*** | (0.249)** | (0.193)** | (0.262) | |
| Negative | 2.097 | 0.264 | 0.424 | 0.14 | 0.285 | 160 |
| | (1.291) | (0.252) | (0.317) | (0.295) | (0.345) | |
| 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 (except for bottom panel which uses Poisson count model) including matching-triplet fixed effects. Dependent variable in third panel is a dummy, indicating if a section reported any case. Column 1 reports standard deviation in parentheses. Column 2-4 report robust standard errors in parentheses. Difference column reports the difference between the CM and NFA coefficients; the standard error is computed using the delta method. Significance: * is significant at the 10% level, ** is significant at the 5% level and *** is significant at the 1% level.

## E.19 Dose-response Models

**Table E.18:** Dose-Response with All Sections in Study Area

| | Total Cases | | Total Cases per Clinic | | Total Cases 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 Cases | | Confirmed Cases per Clinic | | Confirmed Cases 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 |

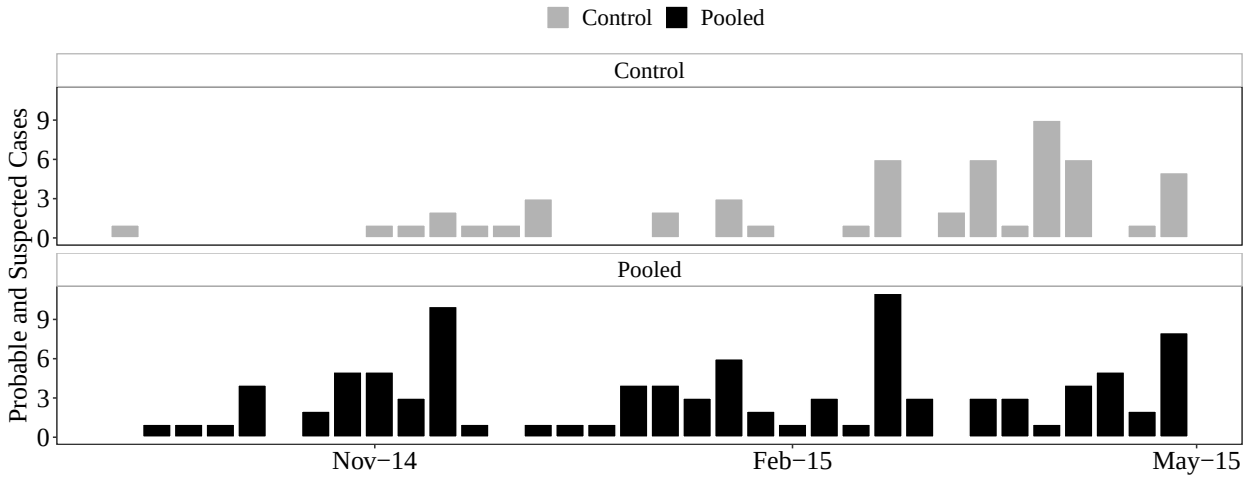
*Notes*: Treatment effects estimated using OLS including matching-triplet and week fixed effects. Robust standard errors, clustered by section, in parentheses. The sample includes all 205 sections that contain at least one clinic from the experimental sample. We regress total reported and confirmed cases on the proportion of clinics within these sections that received either the CM or NFA treatment. Significance: * is significant at the 10% level, ** is significant at the 5% level and *** is significant at the 1% level.

## E.20 Effect on Probable and Suspected Cases

During the study period from September 2014 until April 2015, there are 19 probable and 134 suspected cases. Appendix Figure E.7 plots the time series of probable and suspected Ebola cases by week. 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.7:** Weekly Counts of Probable and Suspected Cases




Bars represent the raw counts of Probable and Suspected Ebola cases by week for control and pooled treatment sections.

**Table E.19:** 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*: Treatment effects estimated using OLS including matching-triplet and week fixed effects. Column 1 reports standard deviation in parentheses. Column 2-4 report robust standard errors, clustered by section, in parentheses. Difference column reports the difference between the CM and NFA coefficients; the standard error is computed using the delta method. Significance: * is significant at the 10% level, ** is significant at the 5% level and *** is significant at the 1% level.

**Table E.20:** 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 | 0.188 | 0.083 | 0.09 | 0.078 | 0.012 | 5,440 |
| | (0.445) | (0.041)** | (0.054)* | (0.049) | (0.062) | |

*Notes*: Treatment effects estimated using OLS including matching-triplet and week fixed effects. Column 1 reports standard deviation in parentheses. Column 2-4 report robust standard errors, clustered by section, in parentheses. Difference column reports the difference between the CM and NFA coefficients; the standard error is computed using the delta method. Difference column reports the difference between the CM and NFA coefficients; the standard error is computed using the delta method. Significance: * is significant at the 10% level, ** is significant at the 5% level and *** is significant at the 1% level.

## E.21 Controlling for Unbalanced Baseline Variables in Ebola Sample

Covariates included based on imbalance reported in Appendix Table E.4: (1) birth in household last year; (2) trust in community; (3) Mende; (4) Temne; (5) highest education level; (6) number of illness or injury cases per household; (7) motorable road; and (8) mobile phone coverage.

**Table E.21:** Effects on Reported Cases Controlling for Unbalanced Baseline Variables

| | Control Mean | Pooled | CM | NFA | Difference | N |
|---|---|---|---|---|---|---|
| Ebola Cases | | | | | | |
| Total | 0.281 | 0.192 | 0.219 | 0.166 | 0.053 | 5,440 |
| | (0.727) | (0.089)** | (0.119)* | (0.1)* | (0.129) | |
| Confirmed | 0.011 | 0.06 | 0.086 | 0.035 | 0.051 | 5,440 |
| | (0.129) | (0.031)* | (0.04)** | (0.032) | (0.039) | |
| Negative | 0.238 | 0.105 | 0.085 | 0.125 | -0.04 | 5,440 |
| | (0.648) | (0.062)* | (0.08) | (0.076)* | (0.094) | |
| IHS(Cases) | | | | | | |
| Total | 0.206 | 0.1 | 0.116 | 0.085 | 0.031 | 5,440 |
| | (0.47) | (0.043)** | (0.058)** | (0.049)* | (0.065) | |
| Confirmed | 0.009 | 0.029 | 0.036 | 0.023 | 0.013 | 5,440 |
| | (0.1) | (0.012)** | (0.016)** | (0.012)* | (0.016) | |
| Negative | 0.179 | 0.066 | 0.063 | 0.068 | -0.005 | 5,440 |
| | (0.433) | (0.035)* | (0.046) | (0.043) | (0.053) | |

*Notes*: Treatment effects estimated using OLS including matching-triplet and week fixed effects. Column 1 reports standard deviation in parentheses. Column 2-4 report robust standard errors, clustered by section, in parentheses. Difference column reports the difference between the CM and NFA coefficients; the standard error is computed using the delta method. Covariates included (all measured as averages): (1) birth in household last year; (2) trust in community; (3) Mende; (4) Temne; (5) highest education level; (6) number of illness or injury cases per household; (7) motorable road; (8) mobile phone coverage, (9) trust in VHC, and (10) prominent village member in household. Significance: * is significant at the 10% level, ** is significant at the 5% level and *** is significant at the 1% level.

## E.22 Bounding spillover effects

**Table E.22:** 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*: Treatment effects estimated using OLS including matching-triplet and week fixed effects. Robust standard errors, clustered by section, in parentheses. Bordering controls is a count of the number of contiguous control sections (mean: 1.08). Bordering controls population measures the population of these contiguous control sections in 1000s (mean: 3.64). Significance: * is significant at the 10% level, ** is significant at the 5% level and *** is significant at the 1% level.

To consider spillovers based on geographic distance, we 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 Appendix Table E.23 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.

Spillovers could occur via road networks. Using data from Open Street Map, we count the number of roads and paths from each section that intersect any other control section in the sample (see Appendix Figure E.8). Appendix Table E.24 interacts this variable with our treatment indicator and, again, finds no indication that treated sections more connected to control sections via the road network see a larger increase in total cases.

Finally, we consider ethnicity, since information and people may move more easily among areas that are similar along cultural lines. 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. Appendix Table E.25 provides no indication that spillovers occur due to movement of patients between proximate co-ethnic areas.

To test whether imprecision in estimating spillovers may obscure spillovers that bias our results to-

**Table E.23:** 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*: Treatment effects estimated using OLS including matching-triplet and week fixed effects. Robust standard errors, clustered by section, in parentheses. 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 (mean: 22.7). Significance: * is significant at the 10% level, ** is significant at the 5% level and *** is significant at the 1% level.

**Table E.24:** Spillovers through Road Network

| | Total | | IHS(Total) | |
|---|---|---|---|---|
| | Model 1 | Model 2 | Model 3 | Model 4 |
| Pooled | 0.173** | 0.159 | 0.083* | 0.080 |
| | (0.084) | (0.108) | (0.043) | (0.054) |
| Pooled × Connected Controls | | -0.050 | | -0.025 |
| | | (0.074) | | (0.041) |
| Connected Controls | | 0.078 | | 0.037 |
| | | (0.071) | | (0.040) |
| Week | 34 | 34 | 34 | 34 |
| Block | 81 | 81 | 81 | 81 |
| N | 5,440 | 5,440 | 5,440 | 5,440 |

*Notes*: Treatment effects estimated using OLS including matching-triplet and week fixed effects. Robust standard errors, clustered by section, in parentheses. Connected controls counts the number of roads and paths that intersect both section *s* and any other control sections in the sample (mean: 1.78). It proxies for the ease of travel to other control sections. Significance: * is significant at the 10% level, ** is significant at the 5% level and *** is significant at the 1% level.

wards zero, we proceed as follows. We assume that instead of the point estimate, the true spillover effect is given by the "most pessimistic" bound of its 95% confidence interval, i.e. the one that would imply the smallest treatment effect. For example, the pooled treatment effect after taking into account spillovers from

**Table E.25:** Spillovers from Proximate Coethnic Sections

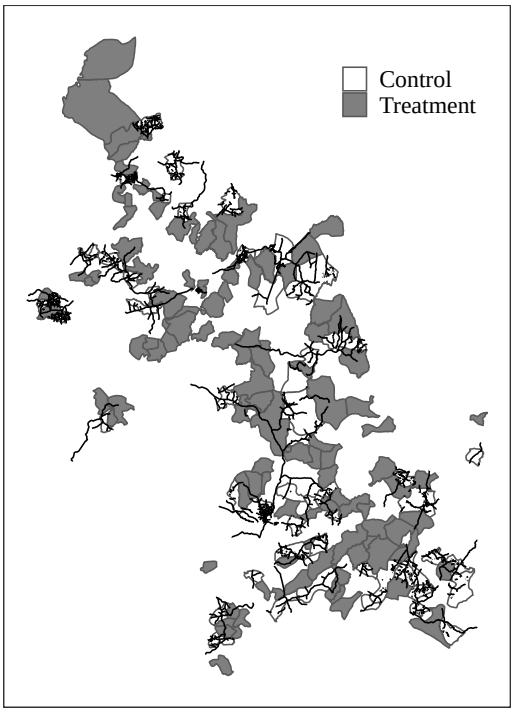
| | Total | | IHS(Total) | |
|---|---|---|---|---|
| | Model 1 | Model 2 | Model 3 | Model 4 |
| Pooled | 0.173** | 0.243** | 0.083* | 0.123** |
| | (0.084) | (0.117) | (0.043) | (0.058) |
| Pooled × Co-ethnic Controls w/in 10 km | | -0.170 | | -0.090 |
| | | (0.111) | | (0.061) |
| Co-ethnic Controls 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*: Treatment effects estimated using OLS including matching-triplet and week fixed effects. Robust standard errors, clustered by section, in parentheses. Co-ethnic controls counts the number of control sections with the same plurality ethnic group (based on the baseline household survey) within a 10 kilometer radius (mean: 0.5). Significance: * is significant at the 10% level, ** is significant at the 5% level and *** is significant at the 1% level.

bordering control sections is 0.360 (Appendix Table E.22, Model 2). The coefficient on the interaction term of treatment and the number of bordering control sections is $-0.134$, with a standard error of 0.116. The most conservative bound of its 95% confidence interval is therefore $-0.134 + 1.96 \times 0.116 = 0.093$. Evaluated at the average number of bordering control sections, 1.08, it is $1.08 \times 0.093 = 0.104$. Under this "pessimistic" assumption about the interaction term, the pooled treatment effect would still be $0.360 - 0.104 = 0.256$. For the analogous model using the IHS (Model 5), we obtain a worst-case pooled treatment effect of $0.179 - 1.08 \times (-0.070 + 1.96 \times 0.060) = 0.128$. Thus, even with very conservative assumptions, the spillover effect is not large enough to overwhelm the estimated treatment effect. We repeat this exercise for all spillover effects, and find a positive adjusted treatment effect—with the exception of Appendix Table E.24, Model 2, which measures spillovers through the road network. Here we find a conservative estimate of $-0.01$, but this emerges from the imprecision with which the spillover coefficient is estimated (standard error: 0.074), not a smaller treatment effect (which has coefficient of $-0.05$). Thus, overall we are able to rule out the possibility that our treatment effects are driven by spillovers.

**Figure E.8:** Roads Intersecting Control Sections

Black Paths: Roads and paths intersecting control sections




Appendix Figure E.8 maps control and (pooled) treatment sections. Black lines indicate roads and paths intersecting control sections.

## E.23 Results for Pre-specified Families in Ebola Sample

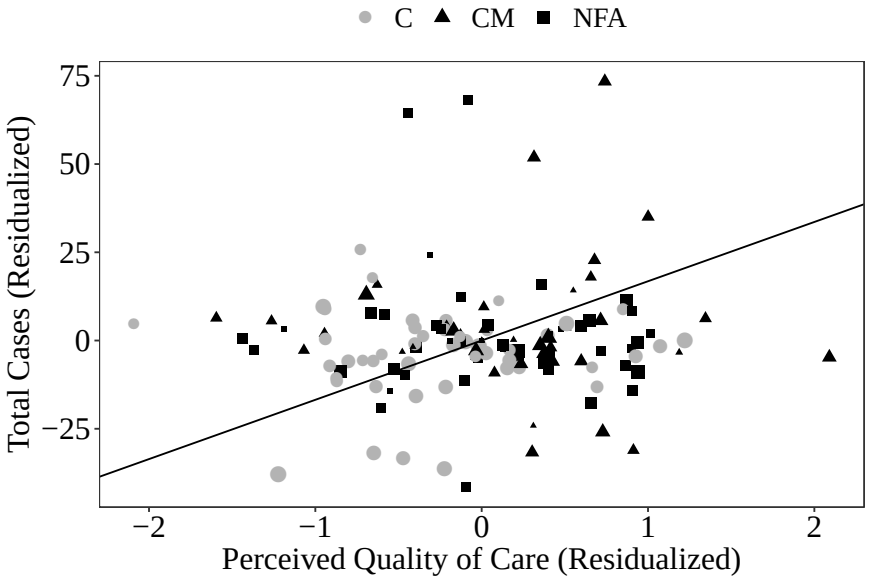
**Table E.26:** Pre-specified Families in Ebola Sample

| | (1) Control Mean | (2) Pooled | (3) CM | (4) NFA | (5) Difference | (6) N |
|---|---|---|---|---|---|---|
| General utilization | 0.000 | 0.074 | 0.128 | 0.036 | 0.092 | 2857 |
| | (1.000) | (0.040)* | (0.049)*** | (0.042) | (0.043)** | |
| | | [0.188] | [0.110] | [0.397] | [0.179] | |
| Maternal utilization index | 0.000 | 0.058 | 0.232 | −0.055 | 0.287 | 595 |
| | (1.000) | (0.086) | (0.112)** | (0.097) | (0.115)** | |
| | | [0.379] | [0.143] | [0.488] | [0.154] | |
| Health outcomes index | 0.000 | 0.034 | 0.085 | −0.006 | 0.090 | 3183 |
| | (1.000) | (0.062) | (0.069) | (0.074) | (0.071) | |
| | | [0.379] | [0.283] | [0.886] | [0.322] | |
| Satisfaction index | 0.000 | 0.105 | 0.101 | 0.107 | −0.006 | 3183 |
| | (1.000) | (0.054)* | (0.067) | (0.060)* | (0.068) | |
| | | [0.188] | [0.235] | [0.144] | [0.593] | |
| Health service delivery index | 0.000 | −0.070 | −0.095 | −0.061 | −0.033 | 1819 |
| | (1.000) | (0.068) | (0.108) | (0.067) | (0.097) | |
| | | [0.234] | [0.373] | [0.397] | [0.534] | |
| Clinic organization and services index | 0.000 | 0.311 | −0.023 | 0.565 | −0.588 | 160 |
| | (1.000) | (0.236) | (0.301) | (0.264)** | (0.315)* | |
| | | [0.205] | [0.459] | [0.144] | [0.214] | |
| CDPE index | 0.000 | 0.264 | 0.232 | 0.288 | −0.057 | 320 |
| | (1.000) | (0.108)** | (0.155) | (0.132)** | (0.184) | |
| | | [0.188] | [0.235] | [0.144] | [0.534] | |
| Contributions to clinic index | 0.000 | 0.173 | 0.292 | 0.080 | 0.212 | 320 |
| | (1.000) | (0.127) | (0.142)** | (0.148) | (0.140) | |
| | | [0.205] | [0.143] | [0.488] | [0.301] | |
| Water and sanitation index | 0.000 | 0.167 | 0.137 | 0.190 | −0.053 | 3183 |
| | (1.000) | (0.086)* | (0.101) | (0.096)* | (0.097) | |
| | | [0.188] | [0.264] | [0.144] | [0.534] | |
| Economic outcomes index | 0.000 | 0.151 | 0.065 | 0.219 | −0.154 | 3183 |
| | (1.000) | (0.087)* | (0.106) | (0.105)** | (0.121) | |
| | | [0.188] | [0.425] | [0.144] | [0.322] | |

*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. Column 1 reports standard deviation in parentheses. Columns 2–5 report robust standard errors, clustered by clinic. Multiple-inference corrected $q$-values that adjust for the false discovery rate within treatment arm across all outcomes in this table 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.24 Visualizing 2SLS Analysis

**Figure E.9:** Visualizing 2SLS Estimate from Table 6




Appendix Figure E.9 uses cross-sectional data from 160 sections. We first residualize both the perceived quality of care index (x-axis) and total cases (y-axis) using matching-triplet and week fixed effects, as well as the baseline value perceived quality of care index. We plot the residualized values, with dots sized according to the weight they receive in the regression. The solid line corresponds to our 2SLS estimate from Table 6 column (1) multiplied by the 34 weeks in our study period. The intuition behind this approach is the following: the (regression-weighted) difference in means between our treatment and control sections along the y-axis corresponds to the impact of the treatments on total cases, i.e., the reduced form of the IV analysis. The (regression-weighted) difference in means between treatment and control sections along the x-axis corresponds to the impact of the treatments on perceived quality of care; i.e., the first stage of the IV analysis. Because the IV coefficient is the reduced form divided by the first stage, the slope of a regression line through this graph is the IV coefficient.

## E.25 Change in perceived quality as predictor of total cases

**Table E.27:** Change in Perceived Quality and Total Ebola Cases in Control Sections

| | *Dependent variable:* | | | |
|---|---|---|---|---|
| | Total Cases | | IHS(Total Cases) | |
| | (1) | (2) | (3) | (4) |
| Change in Perceived Quality (Endline - Baseline) | 0.151 | 0.194* | 0.077 | 0.104* |
| | (0.103) | (0.102) | (0.057) | (0.056) |
| Section Population (1000s) | | 0.064* | | 0.040** |
| | | (0.034) | | (0.019) |
| Observations | 1,836 | 1,836 | 1,836 | 1,836 |

*Notes*: Models estimated using OLS including chiefdom and week fixed effects. Robust standard errors, clustered by section, in parentheses. Sample restricted to control sections. Significance: * is significant at the 10% level; ** at the 5% level; *** at the 1% level.

## E.26 Cost-effectiveness

This section details the cost-effectiveness comparison of CM and NFA relative to Community Care Centers (CCCs). We calculate the average per-clinic cost of implementing both interventions from the financial reports of each of the three implementing NGOs (Plan International, Concern Worldwide, and the International Rescue Committee). We add a five percent administrative overhead based on the structure of the World Bank project that financed these interventions.[A18]

**Table E.28:** Implementation costs by NGO (2013$)

| | | Non-financial awards | | | Community monitoring | | |
|---|---|---|---|---|---|---|---|
| **NGO** | **District(s)** | **Cost** | **Clinics** | **Cost per clinic** | **Cost** | **Clinics** | **Cost per clinic** |
| Plan | Bombali | $ 180,677 | 19 | $ 10,223 | $ 155,358 | 19 | $ 8,790 |
| Concern | Tonkolili | $ 176,304 | 23 | $ 7,665 | $ 136,108 | 23 | $ 5,918 |
| IRC | Bo, Kenema | $ 241,184 | 43 | $ 5,609 | $ 134,409 | 43 | $ 3,126 |
| | | $ 598,165 | 85 | $ 7,037 | $ 425,875 | 85 | $ 5,010 |
| +5% admin overhead | | | | $ 7,389 | | | $ 5,261 |

*Notes*: Authors' calculations based on Plan International's budget and financial reports from Concern Worldwide and the International Rescue Committee, as shared with the World Bank.

On average, the CM intervention cost $7,389 per clinic (in 2013$), and NFA cost $5,261, though there is substantial variation across implementing NGOs. The average, which we can think of as the per clinic cost of the "pooled" intervention, is $6,375.

We calculate the average cost of CCCs from financial documents shared by DFID that show the values of individual contracts awarded to the five implementing NGO partners (Partners in Health, UNICEF, the International Rescue Committee, Oxfam GB, and Plan International) to establish a range of Ebola treatment facilities, including CCCs, Ebola Treatment Units (ETUs), and Ebola Holding Centers (EHCs). We use the contract data for UNICEF, which primarily financed 8-bed facilities, consistent with the WHO definition of a CCC.[A19] The average cost per facility is $707,274 in 2013$,[A20] but this should be seen as a lower-bound estimate, as the per-facility cost varied widely across NGOs and aggregate project spending data from DFID suggests the cost could have been as high as $1.58 million.[A21]

To benchmark the cost-effectiveness of CM and NFA against CCCs, we compare the relative costs

[A18] The interventions were financed as part of a $20m World Bank project loan to the Government of Sierra Leone under the Decentralized Service Delivery Project, which covered several project activities including "social accountability" Makori (2012). The loan allowed for $1m administrative costs, or five percent of the total project budget.

[A19] The World Health Organization defines Community Care Centers (CCCs) as "small facilities (10 beds maximum), located within the community and run by community health workers. CCCs provide isolation facilities for Ebola patients in order to prevent further transmission of the virus within their households and communities. People with Ebola virus can also receive basic curative and palliative care in these centres in an environment supported by their family and communities" WHO (2014b).

[A20] UNICEF financed 44 CCCs for £20.40m in 2014, which gives a per-facility cost of £463,636. To convert to 2013$, we use the 2013-14 UK inflation rate of 2.57% and the 2013 USD/GBP exchange rate of 1.5647: £463,636 x $/£1.5647/1.0257 = $707,274

[A21] DFID's Development Tracker database, which tracks aggregate project disbursements for all DFID-financed projects, shows that the "Ebola Care Units in Sierra Leone" project (IATI # GB-1-204896) disbursed £42,398,846 in 2014, ultimately financing 41 CCCs according to data from the UN Mission for Ebola Emergency Response (UNMEER) Christensen et al.. This implies a per-facility cost of $1.58 million in 2013$. https://devtracker.dfid.gov.uk/projects/GB-1-204896

of the pooled intervention and CCCs to their relative benefits, for which we use the point estimates of the treatment effects on reporting of total Ebola cases, including those in which patients turn out to test negative. Reporting of total possible cases facilitates testing and treatment, ultimately helping to contain the spread of the epidemic. In addition, we also use treatment effects for confirmed cases, as reporting of infected patients may be particularly beneficial for targeting treatment and containment efforts.

We compare the treatment effect from our pooled specification (3) to the equivalent treatment effects of CCCs on increased reporting of Ebola cases from Christensen et al. (2020).[A22] At a cost of $6,375, sections in treatment areas saw 0.173 additional cases per section-week, of which 0.059 were confirmed to be Ebola cases. Over the 34-week period, this equals reporting of 5.88 additional cases of which 2 were confirmed cases per section.

In comparison, at a cost of $707,274, sections with CCCs saw 0.544 additional cases per section-week, of which 0.129 were confirmed to be Ebola (Christensen et al. 2020, Table 1). Over 34 weeks, this equals 18.50 additional cases and 4.39 confirmed cases per section.

To assess cost-effectiveness, we have to answer the following question: Is it more cost-effective to make small, but potentially crucial, preemptive investments in the public health system now, knowing that it will cost more to respond effectively later once an epidemic is underway? The answer depends on (i) comparing the cost of each intervention relative to the gain in Ebola reporting, and (ii) the *ex ante* probability of an outbreak.

We first compare the relative costs listed above to the gains in reporting (total and confirmed cases), using the point estimates of the treatment effects: if $C_{pooled}/C_{CCC} < \hat{\beta}_{pooled}/\hat{\beta}_{CCC}$ the pooled intervention is more cost-effective, with an implied rate of return $\mu_{CM} = \left(\hat{\beta}_{pooled} \times C_{CCC}\right) / \left(\hat{\beta}_{CCC} \times C_{pooled}\right)$. These are listed below in column 3 of Appendix Table E.29. The rates of return for the pooled intervention are 35.28 and 50.74, respectively—in other words, the pooled intervention appears, at first cut, to be orders of magnitude more cost-effective, reflecting in large part the cost of intervening under emergency conditions to try to contain and treat a widespread epidemic.

**Table E.29:** Cost effectiveness of pooled intervention vs CCCs

| **Outcome** | $\beta$(Pooled) | $\beta$(CCC) | $\mu$ | Cost-effective | $p$(Epidemic) |
|---|---|---|---|---|---|
| **Total cases** | 0.173** | 0.544*** | 35.28 | Pooled | 2.83% |
| **Confirmed cases** | 0.059** | 0.129** | 50.74 | Pooled | 1.97% |

*Notes:* Columns 1-2 report treatment effects from Table 3. Column 3 shows the ratio of the treatment effect of the pooled intervention to CCCs, compared to the respective costs of each treatment. Column 4 lists the more cost-effective treatment, CCC or the pooled treatment. Column 5 lists the implied probability for an epidemic that would make the pooled treatment more cost-effective.

However, epidemic outbreak is a probabilistic event that would only be realized after preemptive investments are made. The expected rate of return $E(\mu)$ of intervening now is therefore $\mu$ multiplied by the probability of an epidemic, $p(Epidemic)$; i.e.—if $\mu \times p(E) > 1$ (or equivalently, $p(E) > 1/\mu$ then inter-

[A22] This approach assumes that the marginal social benefit of each outcome and the marginal effect of the treatment on each outcome are both linear; in reality it is likely that there are decreasing returns to both interventions, implying that the benefits to scaling up CM and NFA early on might be higher.

vening now is more cost-effective. For the two outcomes we examine, preemptive investments are more cost-effective than *ex post* interventions under emergency conditions, if there is a 2% or greater probability of an epidemic occurring in any given year. Simulated likelihoods using historical data on pandemics find that the annualized likelihood of another Ebola pandemic of the same scale as 2014–15 is similar at about 1.5% Stephenson et al. (2019).[A23] This suggests that preemptive investments are a worthy investment, not just for their immediate effects, but also as cost-effective ways of preparing for future outbreaks.

[A23] In their updated database, Stephenson et al. (2020) account for the trend of increasing frequency and severity of events and find a significantly higher annualized likelihood of 7.5%. (Based on email correspondence with N. Stephenson, May 28, 2020.)