NBER WORKING PAPER SERIES

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

Darin Christensen
Oeindrila Dube
Johannes Haushofer
Bilal Siddiqi
Maarten J. Voors

Building Resilient Health Systems: Experimental Evidence from Sierra Leone and the 2014 Ebola Outbreak
Darin Christensen, Oeindrila Dube, Johannes Haushofer, Bilal Siddiqi, and Maarten J. Voors
NBER Working Paper No. 27364
June 2020
JEL No. I18,J33,M52,O15

ABSTRACT

Underuse of health systems and a lack of confidence in their quality contribute to high rates of mortality in the developing world. How individuals perceive health systems may be especially critical during epidemics, when they choose whether to cooperate with response efforts and frontline health workers. Can improving the perceived quality of healthcare promote community health and ultimately, help to contain epidemics? We leverage a field experiment to answer this question in the context of Sierra Leone and the 2014 West Africa Ebola crisis. Two years before the outbreak, we randomly assigned two accountability interventions to government-run health clinics – one focused on community monitoring and the other conferred non-financial awards to clinic staff. These interventions delivered immediate benefits under "normal" conditions. Even prior to the Ebola crisis, both interventions increased clinic utilization and satisfaction with healthcare, and 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 significantly reduced Ebola-related deaths. Evidence on mechanisms suggests that the interventions improved confidence in the health system, encouraging patients to report Ebola symptoms and receive medical care. These 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 crises that emerge over the longer run.

Darin Christensen
UCLA Luskin School of Public Affairs
337 Charles Young Drive East
Los Angeles, CA 90095
darinc@luskin.ucla.edu

Oeindrila Dube
University of Chicago
Harris School of Public Policy
1307 E 60th St
Chicago, IL 60637
and NBER
odube@uchicago.edu

Johannes Haushofer
Woodrow Wilson School
Princeton University
427 Peretsman-Scully Hall
Princeton, NJ 08540
and Busara Center for Behavioral
Economics, Nairobi, Kenya
and also NBER
haushofer@princeton.edu/

Bilal Siddiqi
University of California, Berkeley
Center for Effective Global Action
714B University Hall
Berkeley, CA 94720
bilal.siddiqi@berkeley.edu

Maarten J. Voors
Development Economics Group
Wageningen University
Hollandseweg 1, 6706 KN, Wageningen
The Netherlands
maarten.voors@wur.nl

1. Introduction

Over 8 million people die annually in low- and middle-income countries from treatable conditions, generating significant human suffering, and \$6 trillion in economic losses in 2015 alone (Kruk et al. 2018). These deaths are especially tragic: health services are often available at little to no cost, and yet services remain underutilized due in part to the low perceived quality of local healthcare (Dupas 2011; Banerjee et al. 2004; Das et al. 2016). In 2018 and across 12 countries, more than half of patients surveyed by the Lancet Global Health Commission report that they did not seek necessary medical care in the preceding year because they lacked confidence in their health system (Kruk et al. 2018). Lack of public confidence in health systems is also a major barrier to containing epidemics, which require the public to seek testing and comply with isolation and treatment directives. As evidenced by the 2019 novel coronavirus (COVID-19) outbreak, the 2018–19 outbreak of Zika virus, and Ebola outbreaks in the Democratic Republic of Congo (since 2018) and in West Africa over 2014–2016, epidemics and pandemics are a present and ongoing threat with devastating local and global effects.

How can health systems address both persistent maladies and emergent crises? We address this question in the context of Sierra Leone, a country whose chronic health problems were compounded by the West Africa Ebola epidemic. In September 2014, the World Health Organization (WHO) described West Africa's Ebola epidemic as "the most severe acute public health emergency seen in modern times. Never before in recorded history has a biosafety level four pathogen infected so many people, so quickly, over such a broad geographic area, for so long" (WHO 2014a). At that point, less than 7,000 individuals had been infected. By the end of the crisis in early 2016, the Centers for Disease Control and Prevention (CDC) estimated more than 28,000 confirmed, suspected, or probable cases (CDC 2019). Sierra Leone accounts for roughly half of those cases and just under 4,000 deaths.

As with other infectious diseases, Ebola containment efforts emphasize early isolation and treatment. Ebola spreads through the transmission of infected bodily fluids; reducing the reproduction rate thus requires "reducing the time between when people first show symptoms and are isolated" (Whitty et al. 2014). Seeking treatment not only prevents transmission, but also increases the patient's survival prospects: Garske et al. (2017) report case fatality rates of over 90 percent for patients in Sierra Leone with no or unknown hospitalization status, but that rate drops to less than 60 percent for patients admitted to Ebola treatment units or holding centers (see also Waxman et al. 2017).[1]

Yet, fears about sub-standard care and a lack of confidence in health workers and the health system

---

[1] An important caveat is that case fatality rates are difficult to estimate precisely, given that some cases may not be observable, particularly if they do not result in hospitalization or death. Nonetheless, other estimates also suggest that Ebola, untreated, exhibits high fatality rates. For example, Forna et al. (2019) estimate a fatality rate of 89 percent.

deterred symptomatic patients in Sierra Leone from visiting clinics: "Local communities were suspicious of efforts to test, treat, and isolate patients with Ebola symptoms and engaged in practices of hiding sick family members, running away from local communities, or attempting to manage the course of Ebola within local households and communities" (Abramowitz et al. 2016). Thus, under-utilization of health services, reflecting low perceived quality of care, is thought to have been a major obstacle to early treatment and containment of the Ebola epidemic in West Africa.

How can we improve confidence in health systems to boost their resilience to crises? The Lancet Global Health Commission focuses attention on two groups: patients, who must be "accountability agents, able to hold health system actors to account"; and providers, as "demotivated providers cannot contribute to a high-quality health system...[Resilience] requires accountable leaders who respect and motivate their frontline staff" (Kruk et al. 2018).

These recommendations echo claims in personnel economics about how to motivate difficult-to-monitor frontline workers (Finan et al. 2017). While the principal-agent model makes clear how financial incentives can induce effort, performance pay may not always be feasible under resource constraints. Moreover, if frontline health workers perceive intrinsic benefits from their work, financial rewards may crowd out intrinsic motivation (Bénabou and Tirole 2003; Besley and Ghatak 2005; Dixit et al. 2002). Fortunately, non-financial approaches to motivating workers also exist. Organizations can improve accountability and the performance of providers by harnessing social incentives that arise from interactions between providers and clients, or among providers themselves (Ashraf and Bandiera 2018). For example, empowering citizens with monitoring mechanisms to hold providers accountable can improve effort (Mansuri and Rao 2003). Engendering competition among health workers is an alternative approach (Besley and Ghatak 2005).

We test these ideas through a large-scale field experiment implemented just before the Ebola outbreak in Sierra Leone.[2] In partnership with the Government of Sierra Leone (GoSL) and three international NGOs, we randomly assigned two interventions (and control) to 254 government-run health clinics.[3] The timing of the experiment, occurring just before the Ebola epidemic, affords us a unique opportunity to measure the effects of these interventions under both business-as-usual and crisis conditions. Unlike past studies, we can 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.

The first intervention, community monitoring (CM), promotes social accountability by providing patients with information and a public forum to monitor frontline health providers. Modeled on a program evaluated by Björkman and Svensson (2009), the intervention creates scorecards ranking local health services; convenes interface meetings between community members and clinic staff to discuss these ratings

---

[2] Endline surveying for our evaluation concluded in June 2013; the first Ebola case was reported in May 2014 (see Figure 1).

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

and to develop "joint action plans" to improve service delivery; and follows up with meetings to monitor progress after one, three, and nine months. The second intervention motivates clinic staff through a competition that provides non-financial awards (NFA) to both the best and most improved clinic in each district. The competition is advertised to clinic staff, who are also revisited three times during implementation to sustain interest. The winners receive a letter of commendation from the district government and a plaque or wall clock for the clinic. The interventions do not provide physical inputs to clinics; rather, they attempt to empower patients to demand, and motivate clinic staff to supply, higher quality care despite resource constraints.

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

Prior to the crisis, we find that both interventions improve the perceived quality of care and confidence in the health system, as reflected in individuals' satisfaction with health workers and their increased utilization of clinics. Facing a health need, individuals living in treatment areas are seven percent more likely to seek care in a government-run clinic, and satisfaction increases by 0.1 standard deviations. These effects are similar across both the CM and NFA arms. In addition, we find that CM increases clinic utilization by pregnant women and mothers—for example, the probability of delivery in a facility increases by 11 percent—and generates substantial improvements in child health outcomes. Notably, the likelihood of under-5 death in the household declines by 38 percent, likely due to increased maternal utilization. These effects are similar in magnitude to Björkman and Svensson (2009), who find a 33 percent reduction in under-5 mortality in clinics assigned to community monitoring in Uganda. In our experimental context, neither intervention changes the quantity of services provided, nor adds to clinic resources—consistent with the interventions not providing additional inputs to clinics from the central government.

To examine longer-run effects on outcomes related to the Ebola crisis, we use a de-identified database maintained by the GoSL and the CDC. We count the number of reported cases (including individuals who test negative for Ebola) in a given week and section—a small administrative unit with a median area of 40 square kilometers. This analysis focuses on a subset of 160 sections that contain a single clinic from the original sample, thus making its treatment status unambiguous. Pooling the treatments, we estimate that the interventions increased case reporting substantially, by just over 60 percent. While both treatments generate sizable increases, and we cannot reject the hypothesis that the treatments have equivalent effects, we see a larger increase in total reported cases in CM (above 70 percent). While most reported cases test negative for Ebola, the interventions also increase the number of infected patients reporting: a back-of-the-envelope calculation suggests that the effects on reporting reduces the disease's reproduction rate (R_0) by around 19 percent (Pronyk et al. 2016). We find no evidence that geographic spillovers, i.e. the movement of patients

from control to treatment sections, amplify these effects.

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

We attribute the effects on Ebola reporting to improved perceptions of the quality of care and greater confidence in the government health system. Satisfaction and utilization both increase in the Ebola subsample, and respondents in our endline survey express greater confidence in Western ("white-man") medicine relative to traditional healers, the principal alternatives to government clinics in rural Sierra Leone. We combine our measures of utilization, satisfaction with public health workers, and the relative effectiveness of Western medicine into a perceived quality of care index. Instrumenting that index with our treatment assignment, we find that a one standard deviation change in the perceived quality of care increases Ebola reporting by 0.43 cases per section-week (Kling et al. 2007).

We do not find support for two alternative explanations for increased reporting: increased Ebola transmission, and improved disease surveillance. First, while the true incidence of Ebola is unknown (Enserink 2014), preventing us from observing underlying rates of transmission, we see patterns that are inconsistent with the treatments increasing Ebola incidence. Specifically, all types of cases, including patients with both positive and negative test results, increase in sections with treated clinics, and the ratio of confirmed to total cases does not change with treatment status. 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. In addition, we find no evidence of improved disease surveillance: sections with treated clinics do not host more facilities specializing in Ebola care, and there are no differences in laboratory testing or case workers. If anything, we find that a higher proportion of confirmed cases undergo contact tracing (i.e., the process of identifying recent contacts to flag at-risk individuals) in sections with control clinics.

Our results highlight two important points. First, accountability interventions can improve health systems and health outcomes by increasing the perceived quality of care, and by building confidence in health providers. Second, perceived quality of care and confidence in the health system may be especially important during crises, when individuals face a choice about whether to cooperate with response efforts, such as heeding an evacuation order, honoring a quarantine, or voluntarily reporting for medical testing. Trust in government was shown to be correlated with the decision to utilize care during the Ebola crises of Liberia (Blair et al. 2017; Morse et al. 2016; Tsai et al. 2019) and the Democratic Republic of Congo (Vinck et al. 2019)—claims that mirror recent evidence showing that fear and distrust deter patients from utilizing health systems (Alsan and Wanamaker 2017; Lowes and Montero 2018). Experts also argue that low confidence

in health systems undermines efforts to contain the 2019 outbreak of COVID-19.[4] As such, accountability interventions may confer particularly large benefits during crisis conditions through their effects on reporting and cooperation, which can help reduce mortality and contain the spread of an epidemic.[5] We present experimental evidence that accountability interventions can make health systems more resilient to major disruptions such as epidemics, by improving perceptions of care, increasing patient reporting, and lowering mortality conditional on reporting. Other evaluations of accountability programs may miss this potential benefit given the infrequency of crisis events.

Our findings bear on a larger literature on how to increase accountability among frontline public servants and improve public services. Empowering citizens through community monitoring has been examined in a variety of social sectors, including education (Banerjee et al. 2010; Pradhan et al. 2011; Barr et al. 2012; Andrabi et al. 2018), corruption (Fiala and Premand 2018; Olken 2007), and health. As discussed above, our results align closely with Björkman and Svensson (2009), but are further from the more recent work of Raffler et al. (2019), who also examine community monitoring of health clinics in Uganda but find weak effects. As noted by Raffler et al. (2019), one reason for these contrasting results may be the different baseline health conditions. The Björkman and Svensson (2009) intervention was implemented in 2004-2005, when under-5 mortality in Uganda stood at 117 per 1,000 live births. Analogously, our accountability interventions in Sierra Leone were implemented when under-5 mortality stood at 149.8 per 1,000 live births. In contrast, the more recent study in Uganda implemented community monitoring over 2014–2016, when infant mortality had fallen to 59 per 1,000 live births. It may be more difficult to improve health outcomes when baseline conditions themselves have already improved substantially (Raffler et al. 2019).[6] A recent study of community monitoring in Uttar Pradesh (UP), India also finds substantial improvements in immunization rates and treatment of childhood diarrhea, with modest effects on child nutrition outcomes, and no effects on child mortality (Mohanan et al. 2020). The under-five mortality rate stood at 61.15 in UP when the study was conducted, which reinforces the idea that baseline conditions play a role in moderating the effects of community monitoring.

The effects we observe are also consistent with previous work finding that social incentives, such as non-financial awards, can be used to boost worker performance in mission-oriented settings (Ashraf et al. 2014), as well as other experimental settings (Kosfeld and Neckermann 2011; Ball et al. 2001) and

---

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

[5] Our work builds on a literature that examines how other types of individual behavior change in response to epidemics (Agüero and Beleche 2017; Bandiera et al. 2019; Lautharte Junior and Rasul 2019), and how government responds differently when faced with these types of crises (Maffioli 2018).

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

the private sector (Markham et al. 2002). In contrast to these studies, Glewwe et al. (2010) report more narrow gains associated with non-financial awards allocated to teachers who were awarded on the basis of student test scores. The gains they document—on exam scores linked to the incentives, specifically—highlight distortions that can arise when service providers responsible for multiple tasks learn the formula for allocating awards. For this reason, we did not disclose the specific metrics used to assess clinics in our non-financial awards intervention.

The interventions we analyze harness social accountability and competition between providers to improve interactions between providers and the communities they serve. Other related approaches to improving health service delivery have examined the effects of community health workers (Björkman Nyqvist et al. 2019); financial incentives (Miller et al. 2012; Olken et al. 2014; Singh and Mitra 2017); technological monitoring combined with financial incentives (Banerjee et al. 2008); and social incentives among patients, for example, through social signaling (Karing 2019). Our paper is also related to a larger literature that examines various approaches to improving health outcomes in the developing world. A thorough review of this literature can be found in Dupas (2011) and Dupas and Miguel (2017).

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

2. Healthcare 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 five mortality stood at 162.8 deaths per 1,000 live births. 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 healthcare is provided primarily through government-run clinics and hospitals; private and NGO-sponsored facilities are scant (Denney and Mallett 2014). These facilities operate alongside traditional birth attendants and healers. Government clinics, referred to as peripheral health units (PHUs), come

in three types: maternal and child health posts (MCHPs), community health posts (CHPs), and community health centers (CHCs). MCHPs and CHPs are primary health facilities—the first points of contact for patients in towns and villages—that each serve populations of 500 to 10,000 (UNICEF 2014).

In an effort to reduce child and maternal mortality, the GoSL launched a free healthcare initiative in 2010, removing fees for pregnant and lactating women and children under five years old. The policy simultaneously increased pay for government healthcare workers; at the time, 30–50 percent of staff did not receive a government wage and instead relied on charging illegal fees or inflated drug prices and accepting in-kind contributions from the communities they served.

PHUs 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 PHUs were closed.[7] Levy et al. (2015) report that "early assessments [from October 2014] found that many [Ebola] patients were first seeking care at local PHUs." Concerned that these PHUs lacked the training and equipment to properly isolate and care for Ebola patients, PHU staff were rapidly trained on infection prevention and control and outfitted with personal protective equipment. By early December 2014, 81 percent of healthcare workers in Sierra Leone had received training (see Table E.3); by late December 2014, training had reached 98 percent (Levy et al. 2015).[8] In addition to screening and providing "no-touch" treatment for dehydration and fever, a case study in Kenema District found that PHU staff (specifically, community health workers) were also engaged in social mobilization 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 PHUs 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 healthcare workers. Absent incentives tied to service delivery, the government worried that nurses would miss work or continue to charge inflated drug prices—barriers to care that the free healthcare initiative intended to eliminate.

The GoSL contracted with three international NGOs to implement two accountability interventions

---

[7] 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 and, thus, expect negligible (and, if anything, balanced) impacts on clinic access.

[8] Unfortunately, only aggregate data is available on the rollout of training, so we cannot date when individual PHUs were covered. Given the speed at which near universal training was achieved, all clinics in our study area must have received training around the same time.

in 170 clinics across four districts in the country. Plan International worked in Bombali district, Concern Worldwide in Tonkolili district, and the International Rescue Committee worked in Bo and Kenema districts.[9] The four districts bisect Sierra Leone from north to south (see Figure A.2(a)) and cover 15 percent of Sierra Leone's population.

2.2.1 Community Monitoring

The government modeled the community monitoring intervention on Björkman and Svensson's (2009) "Power to the People" approach implemented in Uganda in 2005. The intervention attempts to mobilize "client power," empowering patients with the information and forum to demand accountability from front-line 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 relevant questions in the baseline household survey 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 experiences (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 compact that specified the actions each party would take to improve healthcare. Facilitators worked with both sides to specify a time-frame and assign a responsible "point person" for each component of the compact, and meetings concluded with the signing of the compact by community and clinic representatives.[10] Several

---

[9] Implementation by multiple international NGOs with broad development portfolios suggests that the interventions did not require a local implementer or one specialized in healthcare or social accountability.

[10] The research team randomly selected and observed half of these initial interface meetings. The vast majority of facilitators adhered to meeting protocols. Meetings typically lasted 3–4 hours, and average meeting attendance ranged from 51.9 people in

of the most common problems cited in the compacts 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 encourage use of clinic services and seek prompt and early treatment. After the meeting, facilitators left a copy of the compact with the clinic and representatives from each village.

Finally, facilitators held follow-up meetings one, three, and nine months after the initial inter-face meeting to monitor progress on the joint action plans. These meetings included both community members and clinic staff.[11]

2.2.2 Non-financial Awards

The non-financial awards intervention set up district-wide competitions among clinics, and gave awards to recognize staff performance at both the best and most-improved clinics within each district. Eight awards were allocated across the four study districts; and just under 10 percent of the 85 NFA clinics received an award.[12]

Clinics were ranked at baseline and endline, using baseline data collected at clinics and from households residing in a catchment area around clinics. In each district, awards were given to the best-performing clinic, as well as to the clinic showing the most improvement relative to its baseline ranking. The latter award provided an opportunity to staff at clinics that received low baseline rankings, who might have otherwise been demotivated. 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 services that should be free. 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, 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. 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

---

Kenema district to 68.2 in Bombali district and included representatives from the clinic, traditional authorities, and a larger number of community members (with roughly equal representation of men and women).

[11] The research team observed the three-month follow-up meeting at all clinics that were not visited during the initial interface meeting. In the three-month follow-up, average meeting attendance was only slightly lower than attendance at the first interface meetings.

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

audits uncovered ghost patients or manipulated entries in the clinic register.[13]

The competition was announced and extensively advertised at each clinic in the NFA arm. Winners were not announced until after the endline survey to ensure that any changes reflected the effects of the competition and not health workers' responses to winning or losing. Winning clinics received a plaque or wall clock to display inside the clinic at a public ceremony.

The awards were "non-financial" from the government's perspective, as that 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; our aim is to test whether it had any discernible effects on health outcomes or users' experiences.

3. Design and Methods

3.1 Sampling

3.1.1 Full Sample

The districts in our study include 318 MCHPs and CHPs in total. We sampled 254 of these clinics—192 MCHPs and 62 CHPs—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 that lists the towns and villages it serves. Administrative assignment and high travel costs jointly discourage 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.

---

[13] 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.

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

[Figure 1 about here.]

3.1.2 Ebola Sub-Sample

Our data on Ebola cases are aggregated to the section-week. Sections are the smallest administrative unit in Sierra Leone: the median section is under 40 square kilometers and had fewer than 2,500 residents in the 2004 census (see Figure A.2(c) for a map of section boundaries). We aggregate cases to the section-level since this is the smallest administrative unit for which we can confidently geocode cases. We discuss this in Section 3.2.2 and detail our geocoding procedure in Section E.23. Sections in our primary sample typically contain a single clinic, suggesting these boundaries provide a reasonable approximation of the catchment area.

The 254 clinics in our experimental sample fall into 205 sections. Of these 205 sections, 45 include multiple sample clinics. In our primary analysis, we restrict attention to the 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 these 160 sections, 54 are control, 46 CM, and 60 NFA. As a robustness check, we 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. Results are shown in Table E.12 and are similar to our main results.

[Figure 2 about here.]

As only a subset of clinics were sampled for the study, even sections with one sample clinic can contain

more than one total clinic. However, additional non-sample clinics are rare: among the 160 sections in our main Ebola sample, the share of clinics in each section that are in our sample is, on average across sections, 94 percent. Thus, cases can largely be attributed to the experimental unit. Table E.20 shows that the average share is very similar for treatment and control sections: 93.7 percent in treatment and 93.5 percent in control. The shares are also similar for sections where our sample clinic is a CHP (93 percent) or MCHP (96 percent). There is balance across treatment and control sections in the number of additional clinics that are not in our sample, and this number is generally small.

3.2 Data Collection

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

Baseline data collection took place in September 2011, 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 of each village regarding amenities, relations with the clinic, and community development; and third, household surveys which captured health and economic outcomes and health-seeking behavior among other outcomes.

We filed an analysis plan to examine the survey outcomes at the AEA RCT registry.[14] The plan defines ten outcome families: (1) general utilization, (2) maternal utilization, (3) health outcomes, (4) satisfaction, (5) health service delivery, (6) clinic quality, (7) community development and political engagement, (8) contributions to clinics, (9) water and sanitation, and (10) economic outcomes. The sub-components comprising each outcome family were specified in the analysis plan and are listed in Appendix Section B.1.

Each outcome family represents a set of variables aggregated using control group-standardized indices per Kling et al. (2007). To create an index of K outcomes, we first reverse outcomes where necessary such that a higher value indicates better outcomes. We then compute $\widetilde{y}_i = \frac{1}{K} \sum_{}^{K} \left(\frac{y_{ik} - \mu_{0k}}{\sigma_{0k}} \right)$, where μ_{0k} and σ_{0k} are the estimated control-group mean and standard deviation for outcome k in family K. Our estimates for these families represent standard deviation changes relative to the control group.

Following Kling et al. (2007), in case y_{ik} is missing but another sub-component of the family is mea-

---

[14] 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 dataset, or any outcomes from the clinic or community datasets.

sured, 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. In such cases, we do not impute values when estimating treatment effects for individual sub-components. To demonstrate the imputation is innocuous when looking at effects on families, we follow Kling et al. (2004) and Casey et al. (2012) and aggregate treatment effects across the sub-components of each family using seemingly-unrelated regressions (SUR). These results (reported in Appendix Section D) 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.

1. **General utilization** measures the number of episodes in which individuals seek care at a Western-style medical facility, including in response to four of the most common health needs relevant for primary health units—childbirth in the past year, antenatal or postnatal care, vaccination, or any illness or injury, as well as a residual category of any other type of consultation in the past month. 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 Western-style medical facility reflects the decision to seek care at a formal clinic (overwhelmingly a government-run clinic, given the lack of NGO-run or private clinics), rather than visiting a traditional healer or spiritual leader or forgoing any type of care.

2. **Maternal utilization** is measured among women who gave birth in the year before the endline survey. The family includes two outcomes: an index of the number of times a woman sought antenatal care (ANC) or postnatal care (PNC), and an indicator for whether the woman gave birth in a Western-style medical facility.[15]

3. **Health outcomes** are measured at the household-level. The family includes four measures related to child health: under-5 mortality over the past six months; under-5 illnesses over the past month (e.g., malaria or diarrhea); under-2 vaccine completion; and under-5 child wasting, measured using the weight-for-length ratio.[16] 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

---

[15] Some outcomes within families are themselves indices; for these we continue to use the control-group standardized indices described above.

[16] 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.

family's health; satisfaction with public health workers (i.e., clinic staff); and—among households with at least one member utilizing a Western-style medical facility in the last year (approximately half of the sample)—satisfaction with the care they received.[17] 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 quality** includes three clinic-level outcomes. First, we construct an index of clinic service provision that aggregates measures related to organization (e.g., medicines sorted by expiration date and stored in a safe location), the types and frequency of services offered (e.g., family planning), number of staff on duty, and hours clinics are open. Second, we measure the proportion of staff who are aware of the 2010 policy that removed user fees for maternal and under-5 services. Finally, we measure employee satisfaction. The services offered and employee satisfaction are reported in the clinic survey; other measures are based on enumerators' observations.

6. **Health service delivery** is measured among individuals who experience a health episode in the month before the endline survey (for childbirth episodes, recall is over the past six months). The family includes outcomes derived from the household survey, including staff absenteeism and wait times, problems with clinic facilities or staff, satisfaction with services, staff attitude, drug availability, and fees paid.

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.

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.

---

[17] 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.

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

3.2.2 Ebola Case Data

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

We use a de-identified version of the VHF database (where patient names and characteristics have been redacted).[18] While the data were compiled at the National Ebola Response Center in Freetown, data entry occurred at the district-level. Surveillance officers employed (even prior to the crisis) by the Ministry of Health and Sanitation (MoHS) oversaw teams of case investigators charged with following up on suspected cases. Investigators learned about cases from communities, phone calls, active surveillance (e.g., contact tracing), and through walk-ins at health centers (Owada et al. 2016). For each suspected case, they completed a case investigation form (CIF), which included demographic (including district, chiefdom, and village) and health information. A copy of the CIF accompanied blood samples or swabs from corpses, so that lab results could be linked back to cases to confirm or rule out an infection (McNamara et al. 2016). Completed CIFs were brought back to District Ebola Response Centers and entered by data managers in the local VHF database. Each observation in our data represents one of these CIFs. Data from the districts were periodically transmitted to the National Ebola Response Center.

In the 160 sections that are the focus of our Ebola analysis, the VHF includes 2,045 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;[19] and 19 **probable**

---

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

[19] 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

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. Given our interest in reporting, our primary outcome is 'total cases'—the sum across these four case types.

We aggregate cases to sections based on two features of our geocoding procedure (see Appendix Section E.23). 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 or hamlet. Often this will be the name of a central or headquarter town of the section, which are larger, formal administrative units. By using a larger administrative unit, we avoid measurement error that arises from attributing patients to larger towns that actually occur in the surrounding villages. Second, our geocoding procedure matches residences to lists of geolocated placenames. When we use smaller geographic units, these often contain few placenames 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 contains eight geolocated placenames; and 94 percent of sections contain more than one placename.

We test whether treated sections have more or longer placenames, or placenames that are more likely to contain spaces (i.e., two word names), and find no substantial or significant differences (see Table E.28). To check the accuracy of our geocoding protocol, we look at the average number of confirmed cases in sections that hosted major Ebola-specific treatment facilities: Ebola Treatment Units (ETUs), Ebola Holding Centers (EHCs), or Community Care Centers (CCCs). We find that sections with ETUs, the largest facilities, reported 484 cases on average; sections without any of the three facilities averaged just 25.2 total cases (see Table E.2).

Our main dependent variable is the count of total cases aggregated to the section-week. We use the date when a case is first entered in the VHF database to determine the week. Table E.1 presents descriptive statistics for total and confirmed cases. We show robustness to a number of alternative specifications, including an inverse hyperbolic sine (IHS) and log transformation of the counts, a linear probability model, and a Poisson model (see Table E.9). In the same table, we also report a rare-events logit model, as 20 percent of section-weeks have non-zero total case counts. Results are similar to our main results across these specifications.

---

collected; alternatively, administrative issues may have led to tests being overlooked or not entered into the VHF.

3.3 Randomization

3.3.1 Matching and Blocking

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.[20] Within each triplet, we then randomize clinics into control, community monitoring (CM), or non-financial awards (NFA). This results in 84 clinics assigned to control, 85 to CM, and 85 to NFA. Figure 1 illustrates this allocation of treatment and the timing of randomization relative to data collection. We include matched-triplet fixed effects in our reduced-form specifications to account for the blocked randomization.

3.4 Integrity of the Experiment

Table C.1 reports balance across pre-specified covariates. Most variables are balanced across treatment arms. We find that the number of injuries or illnesses reported is lower in both treatment arms relative to control, household size is slightly smaller, and there are fewer households reporting a recent childbirth in CM. 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, are less likely to belong to the Temne ethic group, are less likely to believe what a doctor told them, and have a higher level of educational attainment.[21] As a robustness check, we include unbalanced variables at baseline as additional controls in Tables D.26 and E.27, and obtain results that are similar to our main results.

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

---

[20] We exactly match clinics by district and type (MCHP or CHP) and then select matches based on the Mahalanobis distance between eight indicators specified by the Ministry of Health 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.

[21] 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 from the community survey: "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.

Nearly 95 percent of staff in NFA clinics have heard of the NFA intervention; conditional on knowing about the program, 84 percent report participating, compared to 48 percent among control villages. Staff in CM and control clinics also appear to have heard about the NFA program, albeit at lower rates. Conditional on having heard about NFA, 13 of 84 control clinics report participating. Again, this likely reflects misperception and not implementation failures or contamination: control clinics could not opt into the NFA competition, which required the research team to rank each participating clinic. Moreover, the research staff did not detect any deviations from treatment assignment when monitoring implementation of the NFA treatment.[22] In any instance, had staff in control clinics been motivated by a mistaken belief that they were eligible for an award, this would only attenuate our treatment effects.

3.5   Specifications

3.5.1   Survey Outcomes

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

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. We instead employ an ANCOVA-type model that controls for the village-level average at baseline, estimating:

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

where $y_{ivc,EL}$ is the outcome of household (or individual) i in village v in clinic catchment c at endline (EL). α_b represents the matched-triplet fixed effects. Treatment status, which is randomized across clinics, is denoted by the indicator variables $\mathbb{1}(CM)_c$ and $\mathbb{1}(NFA)_c$. $\overline{Y}_{vc,BL}$ is the village-level average at baseline. When y is a sub-component of an outcome family, we still use the family-level outcome to compute this baseline average.[23] We cluster our standard errors on clinic, the unit of randomization. We also estimate a variant of Equation 1 in which we pool the CM and NFA treatments into one pooled treatment indicator.

---

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

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

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

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

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.[24]

3.5.2 Ebola Case Data

We assess the impact of the CM and NFA interventions on reported cases for the 160 sections in the Ebola sample (described in Section 3.1.2). We observe counts of reported cases in each section in every week from 10 August 2014 to 18 October 2015. 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 in the seven months between May and October 2015.[25]

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 α_b again represents the matched-triplet fixed effects; δ_t are week fixed effects; $s \in \{1,2,\ldots,160\}$ indexes sections; and $t \in \{1,2,\ldots,34\}$ indexes weeks. For panel models, we cluster our standard errors at the section-level, which, in the Ebola sample, coincides with the clinic, the level of randomization.[26]

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).

---

[24] 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.

[25] In Table E.10 we extend the panel back to August 2014 and replicate our primary results from Table 3.

[26] 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 Medium-run Outcomes

We present the main medium-run survey results in Table 1, which shows estimated effects on clinic utilization, health outcomes, and patient satisfaction. These tables present mean-effects indices per Kling et al. (2007); results using SUR are presented in the top rows of the tables in Appendix Section D.[27] These appendix tables also show how individual indicators contribute to the family-level effects.[28] Our tables follow a common format. Column 1 provides the control mean and standard deviation, which are zero and one exactly when looking at family-level mean-effects indices. Column 2 presents the average treatment effect (in standard deviation units) when pooling the treatment arms in a comparison with control. Columns 3 and 4 separately estimate the average treatment effects for CM and NFA, respectively. Column 5 shows the difference between the average treatment 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.

Table 1 shows uniformly positive treatment effects across families for CM. Specifically, the general utilization of clinics increases by 0.11 standard deviation units when we pool the treatments, and the separate effects are similar for CM and NFA. Individuals in the control group recently used a Western-style medical facility for close to one (0.96) health episode (see Table D.1); the treatments increase utilization of Western clinics by about 5 percent. Our utilization measure includes the use of private or NGO-run clinics, in addition to the government-run clinics (PHUs) targeted by the treatments. These non-governmental providers constitute a small share of utilization (3 percent). When we focus attention on the utilization of just government-run clinics in Table D.21, our treatment effects increase to 7.2 and 5.9 percent for CM and NFA respectively, verifying that the interventions boosted utilization in the targeted clinics.[29]

Maternal utilization—the use of clinics for antenatal care, birthing, and postnatal visits—can only be

---

[27] Overall, our results are very similar using either method of aggregation; though the coefficients on satisfaction and the CDPE family attenuate somewhat using SUR.

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

[29] Our analysis plan specified examining utilization of Western clinics and traditional religious healers. However, due to an error in survey design (see footnote A2 in Appendix Section B.1), we do not have complete utilization data for traditional healers for all health episode types. We therefore focus on utilization of Western clinics in the main results. For one specific episode type, namely illness/injury episodes, we do have utilization data for both Western clinics and traditional healers. We examine these outcomes for illness/ injuries alone in Tables D.22 and D.23. This analysis produces similar results to the main results. We also use the number of visits to Western clinics instead of the proportion of visits out of reported health episodes, as 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 (see footnote A1). However, our results are similar if we use the proportion, or an indicator of any visits to seek care at a Western clinic (results available upon request).

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

The health outcomes family also increases by 0.17 standard deviation units with CM. To a large extent, this is driven by significant improvements in child health. As shown in Table D.3, the likelihood that a child under-5 dies in CM falls by 0.015 relative to the control mean of 0.039—a 38.4 percent effect. In addition, child weight-for-length increases by 0.16 z-score units and is significant at the 10% level, though this individual indicator loses significance after FDR adjustments. It is worth noting that the magnitude of these effects are qualitatively similar to those uncovered by Björkman and Svensson (2009) in their study of community monitoring in Uganda. Finally, the effect size for vaccine completion is also large, corresponding to a 10 percent increase in vaccine completion, though the indicator does not reach significance at conventional levels.

[Table 1 about here.]

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

The final row of Table 1 presents impacts on patient satisfaction. Across both treatment arms, patient satisfaction increases by about 0.10 standard deviations, largely driven by increases in respondents' satisfaction with their own health and health workers (see Table D.4). These effects are consistent with the interventions improving the quality of care provided by health workers in ways that are difficult to capture in survey measures.[30] For example, improvements in quality could include the extent to which health providers listen to patients describing symptoms, or the effort they expend on diagnosis and selecting and explaining an appropriate plan of care, which would translate into greater satisfaction with their performance.[31] The results in Table 1 show that these interventions, which were designed to make frontline health

---

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

[31] It is possible that the effects on satisfaction may partly reflect social desirability bias among respondents in the CM arm, where community members and clinic staff were brought together publicly to discuss the state of local healthcare and health services. It is

providers more accountable to users, can generate substantial improvements in the quality of care that patients perceive when visiting a clinic. Both interventions increase general utilization of clinics, a behavior that reveals users' greater confidence in clinics and patients' reported satisfaction with their health and with health workers.

We now turn to other families of outcomes that might help explain increased utilization and, in CM, improved health outcomes.[32] Specifically, in Table 2, we examine effects on inputs and resources available at clinics (clinic quality), as well as the types and quantity of services provided (health service delivery). We also examine community support for the clinic and community development and political engagement (CDPE). A supply-side response—more resources at clinics or a larger menu of services—or groundswell of community support could also draw would-be patients into clinics and improve their outcomes.

[Table 2 about here.]

Yet, we do not observe effects on the resources or services available at the clinic: the clinic quality and health service delivery families do not change significantly in response to either treatment (or jointly). This is perhaps unsurprising given that the CM and NFA interventions do not facilitate clinics' access to supplies or training; the government's interest in the evaluation was understanding how to extract greater effort from health workers under existing budget and logistical constraints. None of the sub-components of clinic quality significantly improve (see Table D.5). And the coefficients are essentially zero or negative for those sub-components of health service delivery related to supply constraints (e.g., availability of drugs, staff not present (see Table D.6).[33] We do, however, find substantively meaningful effects on some sub-components of health service delivery that relate to patients' experiences. For example, we see a 59 percent reduction in unpleasant staff behavior in NFA areas, though the effect is not statistically significant. The coefficient for staff attitude also suggests improvements, though it loses significance after the FDR adjustments.[34] While there are no significant changes in the services offered at clinics, we do see indications of more

---

less clear why respondents in the NFA arm would feel social pressure to report more satisfaction. Since the effect sizes across the arms are similar in magnitude, this suggests that social desirability bias is unlikely to be the primary driver of the estimated effects.

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

[33] We observe a positive effect of NFA on absenteeism in Table D.6. This is likely an artefact of how we specify this measure: we ask respondents "of all the times you visited the clinic in the past month, did you ever find there were no staff present?" An obvious drawback is that an individual who visits the clinic more frequently has more opportunities to find staff absent. Given the treatment effects on general utilization that we report above, it seems likely that such post-treatment bias pushes towards a positive relationship between the interventions and this measure of absenteeism. Fortunately, we also ask whether respondents found staff absent during their last visit to the clinic. Table D.25 shows precise null effects on this outcome. This comparison suggests that the positive effect on absenteeism in Table D.6 likely reflects increases in utilization, not rates of absenteeism.

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

positive interactions between patients and staff. This could help to explain the improvements in satisfaction, particularly in NFA where we do not find changes in health outcomes.[35]

We also find no increase in community support: community members did not spend more time or resources on the clinic or its staff. We do observe improvements in overall community development and political engagement (CDPE). Yet, the sub-components driving this effect seem unlikely to affect health-seeking behavior (see Table D.8). For example, there are no significant effects on the community coming together to address problems collectively. Instead, the effects are largely driven by contributions to local development projects in both CM and NFA communities and small (< 1.5 percent) increases in voter turnout.[36] We also observe improvements in our water and sanitation family when we pool the treatments, driven by NFA communities (see Appendix Table D.9).[37] Finally, we find only weak effects on economic outcomes (see Appendix Table D.10), suggesting that the interventions did not materially affect households in treated communities.

As a robustness check, we control for imbalanced baseline covariates in Table D.26. 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.

Overall, the positive effects on clinic utilization, health, and satisfaction are not driven by "top-down" improvements in the supply of health services or by greater community contributions to clinics. Rather, people use the clinic to a greater degree, report greater satisfaction (potentially due to improvements in the quality of care, such as more positive interactions with staff), and, in CM, see improvements in child health outcomes. These additional CM effects on health outcomes suggest that engaging the community directly may prove especially consequential in generating changes in clinic utilization that produce health benefits.

4.2   Longer-run Effects on Ebola Outcomes

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

---

[35] 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.

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

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

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

[Figure 3 about here.]

We report regression results using Equation 3 in Table 3. In the top panel, the outcomes are the raw counts of total, confirmed, and negative cases. The pooled effect implies a 62 percent increase in the average number of total cases. The effect is smaller and less precisely estimated for NFA ($p = 0.13$), which is consistent with our medium-run results, which also show smaller effects on utilization for NFA clinics, especially in the Ebola sample (see first row of Table 5). Nonetheless, we cannot reject the null hypothesis that the treatments have equivalent effects. We interpret these effects as reflecting more individuals reporting into clinics if they suspected they had been exposed to Ebola, with the aim of getting tested and seeking treatment. Consistent with this interpretation, we observe increases in cases where individuals turn out to test negative.[38] To both improve patient survival and contain the epidemic, it is particularly important that infected patients report. We find large increases in the average number of confirmed cases reporting in treated sections: for every confirmed case in control, we count five confirmed cases in treated sections.

[Table 3 about here.]

We take a number of steps to assess robustness. First, we re-estimate our pooled effect dropping one matched triplet at a time (Figure E.4), dropping each possible pair of matched triplets (Figure E.5), or dropping each week (Figure E.6). Second, we estimate the effects by month to assess whether our results are driven by a particular moment in the crisis. The pooled coefficient is positive across every month, and we find significant effects in October, December, February and April (Table E.4). (We find large and significant effects for CM in October 2014 and April 2015; for NFA, in October and December 2014.) The spread of these effects across various months throughout our sample period verifies that the effects are not driven by any particular period. Third, we conduct a placebo test where we substitute the nearest out-of-sample neighbor for each section. We find no significant effects in Table E.13, alleviating concerns that our treated sections are spatially clustered in areas where reporting is higher for reasons unrelated to treatment (e.g.,

---

[38] The extent of reporting to clinics is reasonable in light of travel conditions to clinics. The average travel time to clinics in our study is 46 minutes (while average travel distance is 3.2 km). In addition, national quarantines, which would otherwise have restricted travel, were typically in place for a few days at a time.

exposure). Finally, we adopt a number of alternative specifications to handle our count data. In the bottom panel of Table 3, we use the inverse hyperbolic sine (IHS) transformation of the counts and find similar effects; the effect of NFA on confirmed cases becomes significant in this specification. When we use the IHS transformation, the coefficient on the pooled treatment implies a 41.2 percent increase (see Bellemare and Wichman 2020). We expand upon this in Table E.9, which presents estimates using a linear probability model, the logged count (adding one to avoid dropping section-weeks with no cases), a Poisson count model, and a rare-events logit model.[39] In the Poisson count model, NFA has significant effects on both confirmed and negative reported cases; the *p*-value for NFA when analyzing total cases just misses a conventional threshold at 0.104.

Probable and suspected cases (which constitute 1 and 6.5 percent of total cases, respectively) are included in our count of total reported cases. However, these cases often do not involve reporting by individuals; their ambiguous status reflects the absence of a definitive lab test (e.g., confirmed or negative). These cases include, for example, deceased individuals with Ebola symptoms. We separately analyze these cases in Table E.7 and find insignificant and negligible treatment effects.[40] Our results in Table 3 are driven by increases in the number of patients that report and receive testing (see Table E.8 which subtracts probable and suspected cases from total cases).[41] Finally, in Table E.12, we examine 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.

We next look at whether the treatments had any effect on patient deaths.[42] We regress the number of deaths in each section-week on the total number of cases reported in the current and previous week and the interaction of that caseload with treatment. We opt for the caseload over the current and previous week, as Ebola deaths typically occur 6 to 16 days after symptom onset. To ease interpretation, in Table 4 we predict the number of deaths in control and treated sections for a two-week caseload of 2, 5, and 10 cases (see Table E.5 for corresponding regression results). We find significant reductions in mortality: in control sections, we estimate 1 patient death for every 4 cases; that drops to 1 death for every 7 cases in treated

---

[39] With the $\log(y+1)$ transformation, the coefficient on the pooled treatment implies a 45 percent increase. We also collapse the data and estimate cross-sectional models (Table E.11). Our coefficients are of the same magnitude, but we lose power and precision; the Poisson count models remain highly significant with only 160 observations.

[40] 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.

[41] We think it is unlikely that these results are driven by improved record keeping in treatment clinics because (as discussed in Section 3.2.2) Ebola case investigators were not employees of the PHUs but rather a separate team of surveillance officers hired at the district level. These investigators did not simply retrieve official written records from clinics—rather they went to the clinics physically, and followed up with community members to identify and learn more about potential cases. In addition, we observe no differential improvements in record keeping between treatment and control clinics in our endline survey (see Table D.24). As a result, differential changes in record keeping are unlikely to be driving these effects.

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

sections— a reduction driven by CM, where there is just over 1 patient death for every 10 cases. One may worry that patients in control simply waited longer to report and, thus, presented with a higher risk of mortality. Yet, we show in Table E.6 that treatment does not reduce the number of days between symptom onset and reporting. Analogous to our medium-run effects, improvements in health outcomes during the Ebola crisis are concentrated in areas receiving the CM intervention.

[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 treated sections—especially those in CM—enjoyed higher survival rates. Qualitative accounts during the crisis suggest that trust in clinic staff encouraged patients to truthfully report their symptoms and adhere to advice regarding treatment. Raven et al. (2018) also report that health worker morale led to more effective care, especially when treating children.[43] Consistent with the medium-run results, these findings suggest that the accountability interventions improved the care received by patients, even under crisis conditions.

Finally, we consider the implications of our estimates on reported Ebola cases 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.24).

4.2.1 Additional Checks within Ebola Sample

Table E.14 reports balance checks for the 160 sections in our Ebola sample. We find negligible differences along most baseline measures relative to levels in control; the imbalance observed in the full sample carries over to this subset. As a robustness check, we obtain similar results when we aggregate unbalanced baseline indicators to the section-level and include these as controls (Table E.27). We cannot check balance on Ebola caseloads prior to the start of the interventions, since they preceded the 2014 Ebola outbreak by two years. However, in Table E.21, we also look at whether treated sections are more exposed to the epidemic due to their proximity to index cases in Guinea or Sierra Leone, or other characteristics such as road density that could facilitate travel. We find that if anything, treated sections are slightly further from index cases, which would make it harder for us to detect more reported Ebola cases. We also find that treated sections

---

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

do not vary systematically in geographic characteristics including road density, the number of rivers, or the ruggedness of terrain.

We look for evidence of spillovers between treated and control sections, particularly indications that patients traveled from control to treated sections. Such reallocation within our sample would amplify our treatment effects on Ebola reporting. Assuming that patients minimize travel costs, spillovers should be largest in treated sections that border (populous) control sections. In Table E.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.[44] Following a similar logic, 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 Figure E.8). 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, information or people may move more easily between areas that are proximate in both geographic and cultural terms. We use the household survey to determine the plurality ethnic group in each section. Sections tend to be homogeneous: in the median section, 95 percent of respondents report the same ethnicity. For each section, we then count the number of control sections with the same plurality ethnic group and within 10 kilometers. Table E.25 provides no indication that spillovers occur due to movement of patients between proximate co-ethnic areas.

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. Details are given in Appendix Section E.19.

Finally, Ebola cases are a relatively rare event, which raises concerns about power, especially when we estimate effects on confirmed cases. To gauge this, we first ask what magnitude of treatment effects we were powered to detect by multiplying the standard errors on the treatment coefficient in our main analyses by 2.8. As pointed out elsewhere (e.g., Haushofer and Shapiro 2016), this approach yields the treatment effect that we had 80 percent power to detect at the 5% level. For example, the standard error on the pooled treatment

_____

[44] We also calculate each clinic's proximity to the next nearest control clinic in the full sample. (We do not have exact coordinates of clinics, and thus geolocate clinics using the centroids of the census enumeration areas that contain the clinics.) In Table E.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.

effect on total Ebola cases is 0.084 (Table 3), implying that we had 80 percent power to detect treatment effects of $0.084 \times 2.8 = 0.24$, which corresponds to $0.24/0.727 = 0.32$ standard deviations. This analysis suggests we had power to detect only moderately sized effects. To further address concerns that the rarity of the dependent variable affects our estimates, we examine the robustness of the results to a rare-events logit model. The results in Table E.9 show that our main estimates are robust to this alternative estimation strategy.

4.2.2 Mechanisms

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 healthcare and utilization, this would help explain increased reporting in treated sections.[45] Using our endline surveys but restricting attention to the 160 clinics in the Ebola sample (see Table D.20 for estimates using our full sample), in Table 5 we estimate treatment effects on general utilization; satisfaction with public health workers; and households' beliefs about the effectiveness of Western medicine relative to traditional or religious remedies, the primary alternatives to government-run clinics in rural Sierra Leone. The effects on general utilization remain positive and significant when we pool the treatments and in CM alone; the effect is attenuated in the NFA arm relative to the full sample (see Table 1). We continue to find positive effects on satisfaction, focusing here on satisfaction with public health workers, which is asked of all households.[46] Both treatment arms generate roughly equivalent increases in satisfaction with public health workers, on the order of a tenth of a standard deviation. Finally, we find sizable (about 0.3 standard deviation) improvements in households' attitudes towards Western medicine, particularly its effectiveness relative to traditional healers or spiritual remedies. While this index is not listed among the outcomes in our analysis plan, its inclusion was motivated by assessments of the Ebola crisis stressing the importance of trust in Western medicine (e.g., Kruk et al. 2015).[47]

[Table 5 about here.]

---

[45] 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.

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

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

We aggregate the three outcomes in Table 5 into a perceived quality of care index and then instrument this index with our pooled treatment indicator to estimate its effect on reported Ebola cases. In Table 6, we report a large first-stage effect; the F-statistic (9.86) approaches the rule-of-thumb for a strong instrument. When we scale our earlier reduced-form result by this first stage, we find that a one standard deviation change in the perceived quality of care corresponds to an increase in weekly case reports of 0.43 cases per section.[48] While this effect appears large, it is consistent with our hypothesized mechanism that factors such as perceived quality of care become critically important in the context of health crises—when fear of sub-standard care might otherwise prevent people from coming into health facilities and reporting their symptoms.[49]

[Table 6 about here.]

As with the full sample, we do not find consistent positive effects for families focused on supply-side variables in the Ebola sample. Pooling the treatments, we see no significant effects on health service delivery or clinic quality (see Table E.15).[50]

4.3   Alternative Explanations of Ebola Effects

4.3.1   Transmission

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

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

---

[48] Our research design does not separately manipulate treatment status and the level of the mediator. Table 6 quantifies the effect of quality of care on reported Ebola cases only under the strong assumption that the entire effect of the instrument (i.e., the pooled treatment) works through changes in the quality of care. This approach is similar to Kling et al. (2007) who explore whether the Moving to Opportunity program affects individual outcomes through its effect on neighborhood poverty.

[49] We find that endogenous changes in the perceived quality of care were also associated with greater reporting of total cases in control sections. In Table E.31, 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 (to account for the fact that we would expect a larger number of cases in larger sections) 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.

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

We find similar effects on confirmed cases from the two treatments, though NFA did not involve direct out-reach to communities. The CM intervention did include community meetings. If such gatherings continued after the intervention and were sites of Ebola transmission, we would expect more infections in treatment areas to originate from contact with individuals outside of the home.[51] For a subset of infected patients, caseworkers engage in contact tracing, identifying and following up with people who may have come into contact with the infected patient. In this process, caseworkers record how these contacts relate to the patient (e.g., neighbor, tenant, brother, grandmother). In the last two columns of Table E.17, we find that patients subject to contact tracing report fewer contacts outside of their nuclear family (i.e., parents, children, siblings) in CM and NFA relative to control; the number of contacts outside of patients' nuclear families does not increase with CM.

By increasing the number of individuals reporting, the treatments could have increased contact between infected and susceptible individuals, raising the risk of nosocomial transmission.[52] To address this possibility, we compare the dates of symptom onset, reporting, and lab testing. Two features of the Ebola virus are important to note: first, Ebola incubates for 2 to 21 days (8–10 on average) before showing symptoms; and second, an individual can only test positive after displaying symptoms. Consequently, symptom onset or lab results in the first two days after a patient reports cannot not reflect infections due to exposure after the patient reports. For 92 percent of confirmed cases, symptom onset occurs prior to reporting. In 99 percent of cases (all but 2 cases), either symptom onset or laboratory testing occurs within two days of reporting, indicating that nearly all confirmed cases we count do not result from infections that occur after the case was reported.[53]

As further evidence against nosocomial transmission in our sample, Fang et al. (2016) report that infections among healthcare workers fall precipitously by September 2014 (the start of our study period), indicating improved awareness and infection control. We continue to find treatment effects in the months after a nationwide effort during November and December 2014 to train healthcare workers in isolation and no-touch treatment (see Table E.4).

In addition, we look at the ratio of confirmed to total cases across treatment and control areas to determine whether the interventions increased the share of infected patients among total cases. This ratio is however undefined when no cases are reported in a section-week. Below, we take a bounding approach, imputing either all ones or all zeros to observations where the ratio is undefined. Imputing all ones assumes

---

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

[52] In related work, Lowes and Montero (2018) argue that colonial health campaigns in sub-Saharan Africa engendered distrust of medicine due, in part, to nosocomial transmission through the reuse of unsanitary needles.

[53] 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.

that, if cases had reported, they would have all been confirmed; imputing all zeros assumes that, if cases had reported, none would have tested positive. Figure E.9(a) plots the average ratio of confirmed to total cases across control and treated sections. Looking at either bound, there is no meaningful difference in these ratios, and the confidence intervals overlap throughout the study period.[54]

In Section E.21, we write down a model to clarify what must be assumed for our results to reflect a change in exposure (as opposed to reporting). For confirmed cases to increase while the share of confirmed to total to remains unchanged, one must conjecture that the treatments dramatically increased reporting by asymptomatic individuals while having negligible effects among those showing possible signs of the virus. This strains credulity: one cannot preemptively test for Ebola, so individuals without symptoms have no reason to report; moreover, qualitative accounts suggest the crisis deterred unexposed individuals from visiting clinics, even when they had other healthcare needs (Elston et al. 2016).

4.3.2 Surveillance

Since treatment clinics were more exposed to three international NGOs and may have had more communication with the government Health Ministry, this raises the possibility that they had greater capacity to reach out to community members and pursue a set of supply-side activities during the crisis that manifest in more individuals reporting. Disease surveillance is challenging in a setting like rural Sierra Leone: if individuals or families want to avoid health workers, they are unlikely to be detected given the difficulties of canvassing sparsely populated regions with limited road networks (Richards et al. 2015; Olu et al. 2016; McNamara et al. 2016). Nonetheless, we look for indications of intensified top-down efforts in treated sections as a possible alternative explanation for the increase in reported cases.

First, contact tracing is central to disease surveillance efforts. In our control sections 59 percent of confirmed cases were subject to contact tracing, compared to just 22 percent in CM and 28 percent in NFA (Table E.17). Second, we use three measures derived from the VHF data (all measured at the section-level): (1) the probability that a case received laboratory testing to confirm or rule out an infection; (2) the average number of days that passed between a case being reported and lab testing; and (3) the number of unique case workers (logged) that entered information into the VHF. We expect these variables to proxy for top-down surveillance efforts during the crisis. In Table E.18, we find no significant differences for these variables across treatment and control.

Second, using data from Sierra Leone's National Ebola Response Center (NERC) and the UN Mission

---

[54] It is possible that the ratio of confirmed to total cases could stay constant if there was an increase in the number of probable and suspected cases. Figure E.9(b) repeats the bounding exercise but uses the ratio of confirmed to confirmed plus negative cases. This exercise delivers the same conclusion, as the number of probable and suspected cases are small and unaffected by treatment (see Table E.7).

for Ebola Emergency Response (UNMEER), we count the number of Ebola-specific treatment facilities in each section (see Section E.2). There were three types of specialized facilities: Ebola Treatment Units (ETU, 32 beds on average), Ebola Holding Centers (EHC, 18 beds on average), and Community Care Centers (CCC, 10 beds on average). Only one ETU falls within our sample, and it is located in a control section; Table E.19 shows no significant difference in the counts—either combined or separate—of EHCs or CCCs.[55]

Our results in Table 3 are robust to dropping the small number of sections that contain one or more of these specialized facilities: when analyzing total reported cases, the coefficient on our pooled treatment indicator increases from 0.173 to 0.177 (se = 0.091) when we drop these sections; for confirmed cases, it changes from 0.59 to 0.55 (se = 0.025) (detailed results available upon request).[56]

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 Table E.3). We find no indications that greater exposure to international partners build capacity for disease surveillance.

4.4 Cost Effectiveness

Our results show that these accountability interventions can not only improve outcomes over the medium run, but can 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 local healthcare, particularly maternal and child care. 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 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 medical facilities and, thus, encourage reporting and early isolation and treatment (Michaels-Strasser et al. 2015). They were widely considered effective and the cheapest among the emergency response centers (as

---

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

[56] 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 Table E.20 we find that treated sections are not significantly further from ETUs, EHCs or CCCs in the NERC data; the distance from NFA sections to the nearest CCCs is shorter when we use the UNMEER data.

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.25 for details on how these costs are calculated).

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 accountability 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 accountability interventions incur costs without contributing to containment. Comparing the ratios of cost and effect sizes implies that the accountability interventions are more cost-effective than CCCs for epidemic events with >2–3 percent probability of occurring (see Appendix Section E.25).[57]

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.

5.   Conclusion

This study shows that accountability interventions, such as community monitoring and non-financial awards, can boost the perceived quality of health care and improve health outcomes in a developing country setting—

---

[57] This statement again focuses solely on epidemic outcomes, setting aside the medium-run benefits, such as clinic utilization and child health that emerge in times without epidemics.

not only during "normal" times, but also during crises. We use a randomized experiment completed less than a year before the Ebola outbreak in Sierra Leone to test the effectiveness of two interventions: one focused on community monitoring of government-run health clinics, and the other on status awards for clinic staff.

In the medium term, we find that both interventions increase patient utilization, satisfaction, and beliefs about the relative effectiveness of Western medicine. This suggests that the accountability interventions improved the perceived quality of healthcare and built confidence in the health system. We see these improvements in confidence translating into immediate gains in critical health outcomes, particularly under community monitoring. Similar to previous work, we find large reductions in under-5 mortality over the medium run (Björkman and Svensson 2009). This result highlights how direct community engagement may be especially powerful for changing utilization patterns in ways that prove consequential for health outcomes.

We also find evidence that improved confidence in the health system led to greater willingness to get tested and seek care during the Ebola epidemic, which struck Sierra Leone approximately two years after the interventions were implemented. The interventions substantially increased reporting of Ebola cases during the crisis—by both patients who tested positive and negative for the virus—while mortality among Ebola patients fell. This is consistent with testing enabling treatment and ultimately reducing the viral reproduction rate and highlights how increased testing can save additional lives in epidemic contexts.

We explore two alternative explanations for why the interventions could have increased reported case counts—by unintentionally increasing exposure to Ebola or enabling more top-down surveillance efforts. We do not find support for either mechanism. Specifically, we find no evidence to suggest that the interventions contributed to transmission at treated clinics or raised the infection rate among patients (i.e., the ratio of confirmed to total cases). Rather, we observe increased reporting by both infected individuals, as well as those who feared they had been exposed but tested negative. We also see no indication of more Ebola-specific treatment facilities, lab resources, or caseworkers in treated areas, suggesting that resources for screening and contact-tracing were not targeted to areas that received the interventions.

The substantial effects we estimate for community monitoring on patient outcomes over the medium run align with the landmark study of Björkman and Svensson (2009) in Uganda and recent work by Mohanan et al. (2020) in India. However, they contrast with Raffler et al. (2019), also set in Uganda, but conducted more recently. This discrepancy raises questions around external validity: under what conditions should we expect to observe substantial treatment effects from accountability interventions? As noted in the introduction and discussed at length by Raffler et al. (2019), one potential explanation is differing baseline health conditions. It may be more difficult to further improve outcomes like child mortality in contexts where these indicators are already fairly good or have recently improved, as in Uganda during Raffler et al.'s (2019) study period. In contrast, baseline health conditions, including child morality, were very poor in Sierra

Leone during our study period.

We also uncover longer-run effects during the Ebola crisis. While the increases in patient utilization over the medium run are modest, the effects on reporting during the Ebola epidemic are substantial. This suggests that even moderate shifts in the perceived quality of care may strengthen resiliency and generate large improvements when crises hit. We know of no comparable experimental results from other contexts; such effects are difficult to capture, as they only emerge with the onset of health crises. It thus remains an important open question whether accountability interventions bolster reporting and resiliency in other places and crises. Related work on trust and Ebola outbreaks in Liberia and the Democratic Republic of Congo points to wider applicability (Blair et al. 2017; Morse et al. 2016; Tsai et al. 2019; Vinck et al. 2019). And our analysis of mechanisms suggests that such effects should manifest where patients otherwise lack confidence in local healthcare. This is a condition that Kruk et al. (2018) find is all too common across low- and middle-income countries. Unfortunately, the COVID-19 outbreak revealed that distrust of public health efforts can undermine disease containment even in high-income countries. While health system capacity is critical—testing and tracing, for example, require mass mobilization of suppliers—building trust among citizens engenders cooperation and may encourage the widespread behavioral change needed to curb contagion. Our results suggest that (even preemptive) investments in confidence-building interventions constitute a promising approach to preparing for future crises.

References

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

Agüero, Jorge M and Trinidad Beleche (2017). "Health shocks and their long-lasting impact on health behaviors: Evidence from the 2009 H1N1 pandemic in Mexico." *Journal of Health Economics* 54, 40–55.

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

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

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

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

Ashraf, Nava, Oriana Bandiera, and B Kelsey Jack (2014). "No margin, no mission? A field experiment on incentives for public service delivery." *Journal of Public Economics* 120, 1–17.

Ball, Sheryl, Catherine Eckel, J Grossman, Philip, and William Zame (2001, 02). "Status in markets." *Quarterly Journal of Economics* 111(1), 161–188.

Bandiera, Oriana, Niklas Buehren, Markus Goldstein, Imran Rasul, and Andrea Smurra (2019). "The Economic Lives of Young Women in the Time of Ebola." *World Bank Policy Research Working Paper*.

Banerjee, Abhijit, Angus Deaton, and Esther Duflo (2004). "Health care delivery in rural Rajasthan." *Economic and Political Weekly* 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." *Unpublished 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, May). "Power to the People: Evidence from a Randomized Field Experiment on Community-Based Monitoring in Uganda." *Quarterly Journal of Economics* 124(2), 735–769.

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

Björkman Nyqvist, Martina, Andrea Guariso, Jakob Svensson, and David Yanagizawa-Drott (2019, July). "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.

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

Casey, Katherine, Rachel Glennerster, and Edward Miguel (2012). "Reshaping institutions: Evidence on aid impacts using a preanalysis plan." *The Quarterly Journal of Economics* 127(4), 1755–1812.

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

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

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

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

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

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

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

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

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

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

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

Forna, Alpha, Pierre Nouvellet, Ilaria Dorigatti, and Christl A Donnelly (2019, 07). "Case Fatality Ratio Estimates for the 2013–2016 West African Ebola Epidemic: Application of Boosted Regression Trees for Imputation." *Clinical Infectious Diseases* 79(1), 128.

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

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

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

Haushofer, Johannes and Jeremy Shapiro (2016). "The short-term impact of unconditional cash transfers to the poor: Experimental evidence from Kenya." *The Quarterly Journal of Economics* 131(4), 1973–2042.

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

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

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

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

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

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

Lautharte Junior, Ildo and Imran Rasul (2019). "The Anatomy of a Public Health Crisis: Household and Health Sector Responses to the Zika Epidemic in Brazil."

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

Lowes, Sara and Eduardo Montero (2018). "The Legacy of Colonial Medicine in Central Africa." *Working Paper* (https://scholar.harvard.edu/slowes/publications/colonial-medicine).

Maffioli, Elisa M (2018). "The Political Economy of Health Epidemics: Evidence from the Ebola Outbreak." *Available at SSRN 3383187*.

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, July). "Innovation to confront Ebola in Sierra Leone: The community-care-centre model." *The Lancet Global Health* 3(7), e361–e362.

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

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.

Mozur, Paul (2020). "China, Desperate to Stop Coronavirus, Turns Neighbor Against Neighbor." *New York Times* (https://www.nytimes.com/2020/02/03/business/china-coronavirus-wuhan-surveillance.html).

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

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

Olu, Olushayo Oluseun, Margaret Lamunu, Miriam Nanyunja, Foday Dafae, Thomas Samba, Noah Sempiira, Fredson Kuti-George, Fikru Zeleke Abebe, Benjamin Sensasi, Alexander Chimbaru, Louisa Ganda, Khoti Gausi, Sonia Gilroy, and James Mugume (2016). "Contact Tracing during an Outbreak of Ebola Virus Disease in the Western Area Districts of Sierra Leone: Lessons for Future Ebola Outbreak Response." *Frontiers in Public Health* 4, 130–9.

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

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

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

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

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

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

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

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

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

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

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

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

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

WHO (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 (n.d.). "Sierra leone: Analytical summary - malaria. http://www.aho.afro.who.int/profiles_information/index.php/Sierra_Leone:Analytical_summary_-_Malaria, last accessed on April 20, 2019.
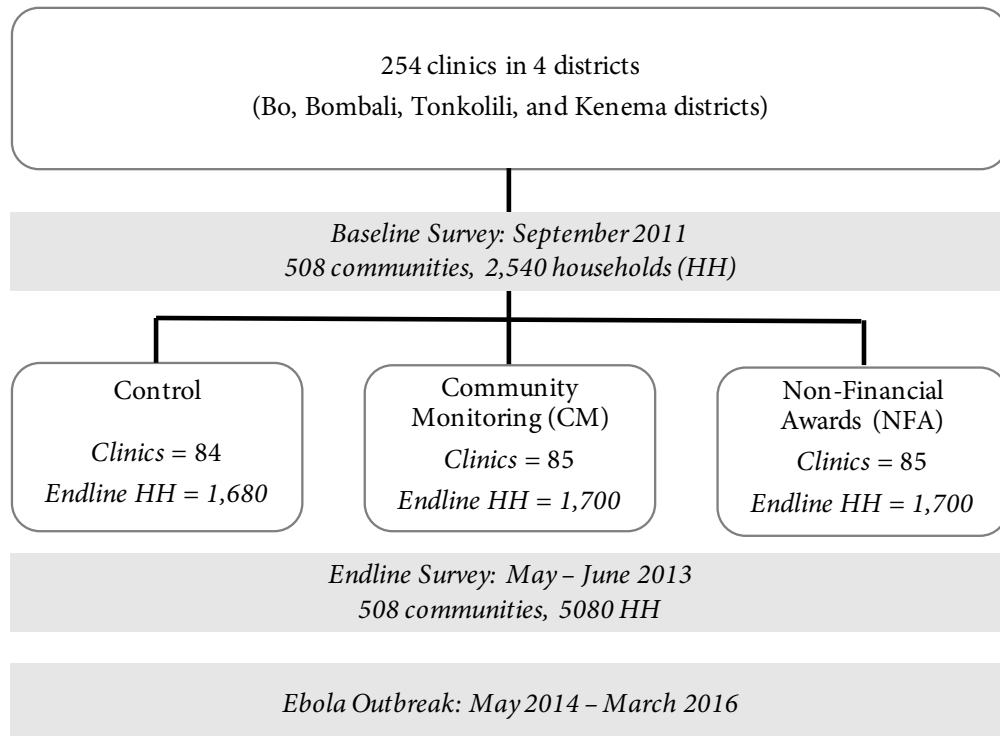
Figure 1: Consort Diagram

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
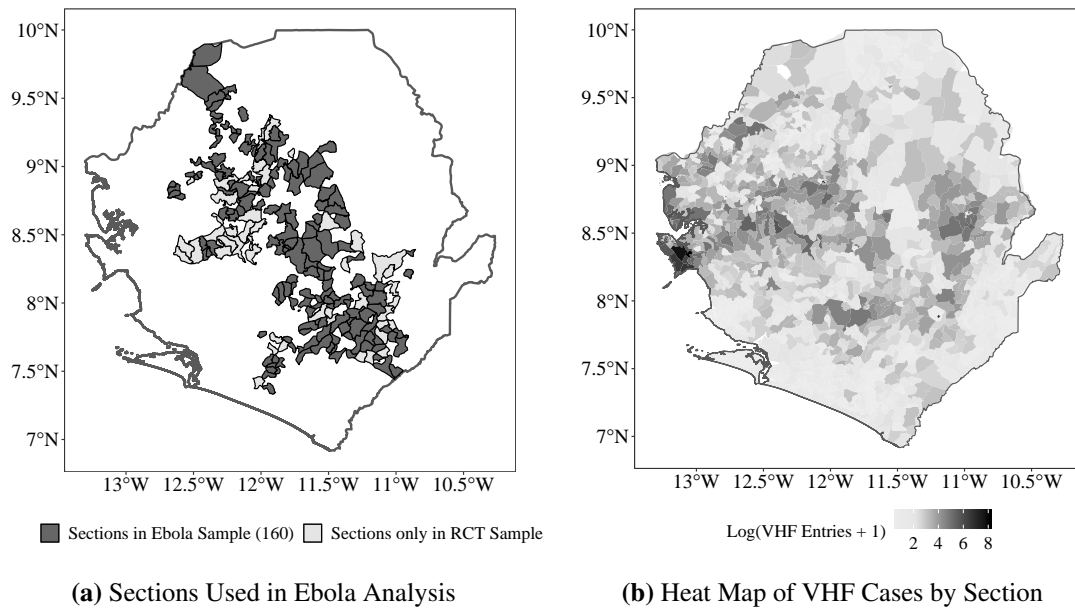
Figure 2(a): map of all sections that contain clinics that were part of the original randomized experiment. The 45 sections in light gray are excluded from the primary Ebola analysis, because they contain more than one clinic from the original RCT. Figure 2(b): the number of entries by section in the 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 Cases by Treatment

(a) Weekly Counts

(b) Cumulative Counts

Figure 3(a) plots the time series of total cases by week; bars represent the raw counts. 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 cases by treatment group.
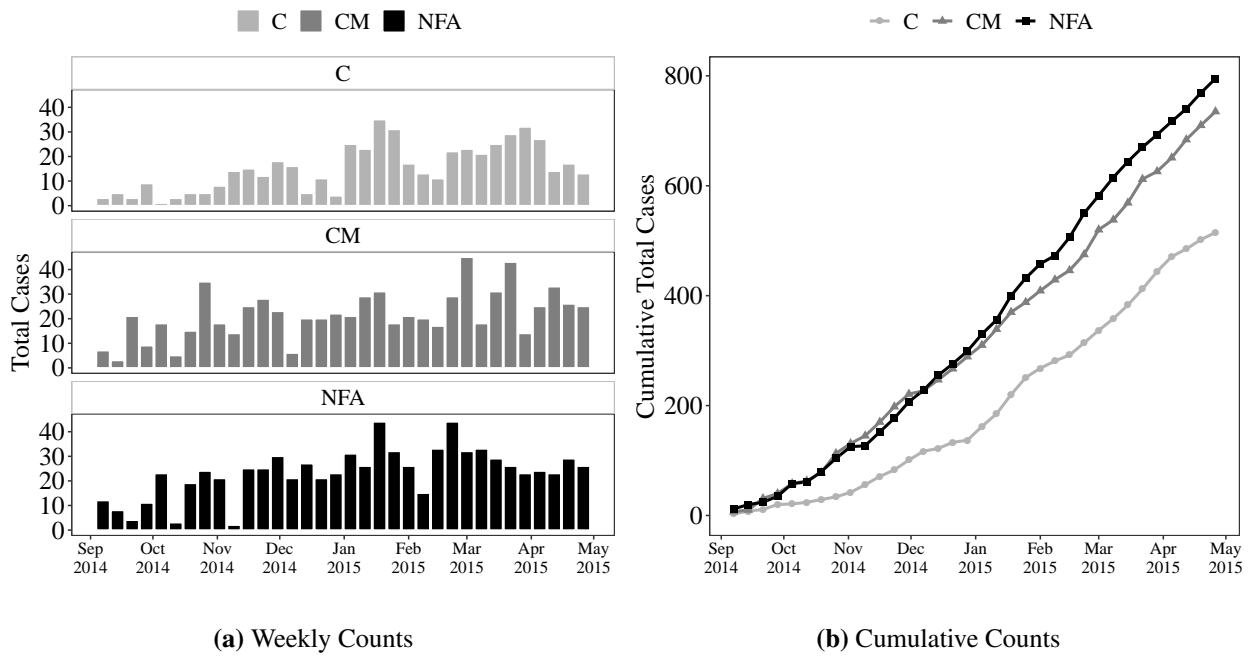
44

Table 1: Utilization, Satisfaction, and Health Outcomes

	(1) Control Mean	(2) Pooled	(3) CM	(4) NFA	(5) Difference	(6) Joint F-test (p)	(7) N
General utilization	0.000 (1.000)	0.112 (0.031)*** [0.005]***	0.126 (0.034)*** [0.003]***	0.099 (0.037)*** [0.032]**	0.026 (0.033)	7.054 (0.001)***	4496
Maternal utilization	0.000 (1.000)	0.061 (0.064) [0.327]	0.175 (0.077)** [0.068]*	−0.043 (0.076) [0.548]	0.218 (0.081)***	4.128 (0.017)**	888
Health outcomes	0.000 (1.000)	0.064 (0.051) [0.265]	0.166 (0.055)*** [0.014]**	−0.039 (0.060) [0.548]	0.205 (0.055)***	8.105 (0.000)***	5053
Satisfaction	0.000 (1.000)	0.101 (0.042)** [0.038]**	0.091 (0.049)* [0.095]*	0.109 (0.049)** [0.048]**	−0.018 (0.048)	2.876 (0.058)*	5052

Notes: Treatment effects are estimated using Missing-Indicator ANCOVA, controlling for the community-level average of the outcome family index at baseline and matching-triplet fixed effects. Robust standard errors, clustered by clinic, are shown in parentheses. Multiple-inference corrected q-values that adjust for the false discovery rate within treatment arm 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. 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 2: Supply-side Measures and Community Support

	(1) Control Mean	(2) Pooled	(3) CM	(4) NFA	(5) Difference	(6) Joint F-test (p)	(7) N
Clinic quality	0.000	0.104	−0.004	0.213	−0.216	0.929	254
	(1.000)	(0.149)	(0.175)	(0.176)	(0.184)	(0.397)	
		[0.395]	[0.649]	[0.237]			
Health service delivery	0.000	0.039	0.070	0.027	0.043	0.507	2877
	(1.000)	(0.059)	(0.071)	(0.062)	(0.059)	(0.603)	
		[0.395]	[0.266]	[0.583]			
Community support	0.000	0.033	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]			
CDPE	0.000	0.231	0.202	0.261	−0.059	3.849	508
	(1.000)	(0.085)***	(0.102)**	(0.101)**	(0.110)	(0.023)**	
		[0.034]**	[0.095]*	[0.032]**			

Notes: Treatment effects are estimated using ANCOVA, controlling for the community-level average of the outcome family index at baseline and matching-triplet fixed effects. Robust standard errors, clustered by clinic, are shown in parentheses. Multiple-inference corrected *q*-values that adjust for the false discovery rate within treatment arm 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. 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 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. 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 bottom panel employs the inverse hyperbolic sine transformation ($\text{IHS}(y) = \log(y + \sqrt{(1 + y^2)})$). Significance: * is significant at the 10% level, ** is significant at the 5% level and *** is significant at the 1% level.

Table 4: Patient Deaths per Section per Week

Total Cases in Last 2 Weeks	Predicted Deaths in Control (1)	Predicted Deaths in Pooled (2)	Difference (3)
2 reported cases	0.49	0.36	0.13
	(0.04)	(0.05)	(0.06)**
5 reported cases	1.23	0.80	0.43
	(0.11)	(0.17)	(0.19)**
10 reported cases	2.45	1.53	0.92
	(0.21)	(0.36)	(0.40)**

Notes: 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 E.5 model 1. 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
General utilization	0.000	0.074	0.128	0.036	0.092	3.704	2857
	(1.000)	(0.040)*	(0.049)***	(0.042)	(0.043)**	(0.027)**	
Satisfaction with public health workers	0.000	0.157	0.157	0.156	0.001	5.116	3149
	(1.000)	(0.049)***	(0.064)**	(0.054)***	(0.064)	(0.007)***	
Relative effectiveness of Western medicine	0.000	0.293	0.340	0.257	0.084	2.867	3183
	(1.000)	(0.123)**	(0.160)**	(0.132)*	(0.154)	(0.060)*	

Notes: Treatment effects are estimated using Missing-Indicator ANCOVA, controlling for the community-level average of the outcome family index at baseline and matching-triplet fixed effects. Robust standard errors, clustered by clinic, are shown in parentheses. The F-test column provides evidence on the joint significance of CM and NFA, with the associated p-value in parentheses. 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 6: Perceived Quality of Care and Ebola Cases

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

Notes: Matching-triplet fixed effects included in both first- and second-stage regressions. Robust standard errors, clustered by section, in parentheses. "Perceived Quality of Care" is an equally weighted average of the variables in Table 5. Significance: * is significant at the 10% level, ** is significant at the 5% level and *** is significant at the 1% level.