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

DARIN CHRISTENSEN
OEINDRILA DUBE
JOHANNES HAUSHOFER
BILAL SIDDIQI
MAARTEN VOORS

Skepticism about the quality of health systems and their consequent underuse are thought to contribute to high rates of mortality in the developing world. The perceived quality of health services may be especially critical during epidemics, when people choose whether to cooperate with response efforts and frontline health workers. Can improving the perceived quality of health care promote community health and ultimately help to contain epidemics? We leverage a field experiment in Sierra Leone to answer this question in the context of the 2014 West African Ebola crisis. Two years before the outbreak, we randomly assigned two interventions to government-run health clinics—one focused on community monitoring, and the other conferred nonfinancial awards to clinic staff. Prior to the Ebola crisis, both interventions increased clinic utilization and patient satisfaction. Community monitoring additionally improved child health, leading to 38% fewer deaths of children under age five. Later, during the crisis, the interventions

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

The Quarterly Journal of Economics (2021), 1145–1198. doi:10.1093/qje/qjaa039.
Advance Access publication on November 20, 2020.

also increased reporting of Ebola cases by 62%, and community monitoring significantly reduced Ebola-related deaths. Evidence on mechanisms suggests that both interventions improved the perceived quality of health care, encouraging patients to report Ebola symptoms and receive medical care. Improvements in health outcomes under community monitoring suggest that these changes partly reflect a rise in the underlying quality of administered care. Overall, our results indicate that promoting accountability not only has the power to improve health systems during normal times, but can also make them more resilient to emergent crises.
JEL Codes: I18, J33, M52, O15.

I. Introduction

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

How can the quality of health care be improved? Do programs that achieve this goal under normal conditions also work when crises hit? We address these questions in the context of Sierra Leone, a country whose chronic health problems were compounded by the West African Ebola crisis. In September 2014, the World Health Organization (WHO) described the epidemic as "the most severe acute public health emergency seen in modern times" (WHO 2014). By the end of the crisis in early 2016, the Centers for Disease Control and Prevention (CDC) estimated more than 28,000 confirmed, suspected, or probable cases, with Sierra Leone accounting for roughly half of those cases and just under 4,000 deaths (CDC 2019).

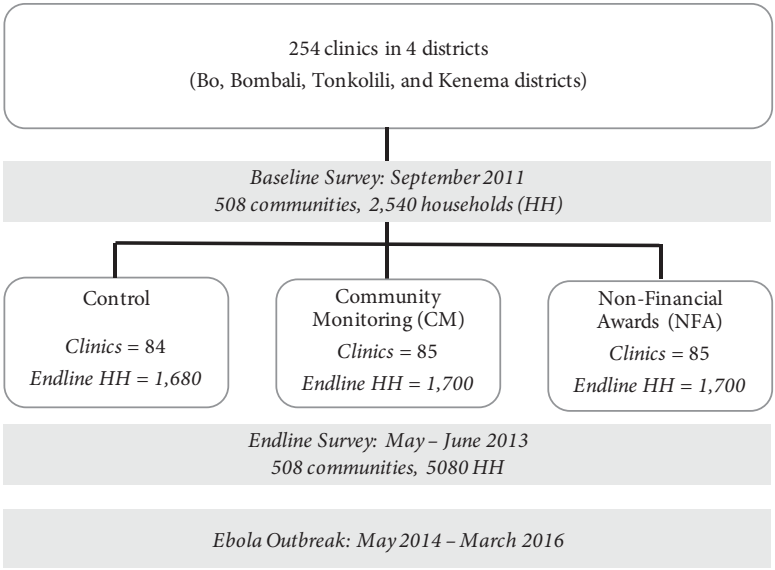


FIGURE I

CONSORT Diagram

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.

Prior to Sierra Leone's Ebola outbreak, we designed a large-scale field experiment to evaluate two programs intended to improve the utilization of government-run clinics and the quality of care delivered at these facilities. The timing of our study enables us to examine the programs' effects both under "normal conditions", and during the ensuing Ebola crisis. Endline surveying concluded in June 2013; the first Ebola case was reported in May 2014 (see Figure I). 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.

We randomly assigned 254 clinics to one of the two interventions or control, in partnership with the Government of Sierra Leone (GoSL) and three international NGOs.[1] The first

1. The interventions were funded by the World Bank and implemented by the NGOs Concern Worldwide, the International Rescue Committee, and Plan

intervention, community monitoring (CM), provided patients with information and a public forum to monitor frontline health workers. Modeled on a program evaluated by Björkman and Svensson (2009), the intervention distributed scorecards to rate local health services and convened meetings between community members and health workers to discuss these ratings and develop "joint action plans" to improve service delivery. The second intervention provided nonfinancial awards (NFAs) to both the best and most-improved clinic in each district. Clinic staff were encouraged to develop action plans, and the winning clinics received wall plaques and letters of commendation from the district government. Neither program provided resources to clinics; rather, they intended to motivate health workers to supply higher-quality care under existing resource constraints. The programs draw on insights from personnel economics about how to motivate difficult-to-monitor frontline workers (Finan, Olken and Pande 2017). One strand of this literature focuses on nonmonetary approaches, recognizing that performance pay may not be financially feasible or could crowd out intrinsic motivation (Dixit et al. 2002; Bénabou and Tirole 2003; Besley and Ghatak 2005). Organizations can improve workers' performance by harnessing social incentives that arise from interactions between providers and clients or among providers themselves (Ashraf and Bandiera 2018): the CM program empowers citizens to monitor providers and sanction those who underperform (Mansuri and Rao 2003), while the NFA program engenders competition among health workers to improve service delivery (Besley and Ghatak 2005).

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

International, with support from GoSL's Decentralization Secretariat and the Ministry of Health and Sanitation.

patient satisfaction, including satisfaction with the performance of health workers. These results are again consistent with improvements in the perceived quality of care provided by staff at program clinics.

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

We assess the effects of these programs during the ensuing Ebola epidemic. Ebola containment efforts emphasize early isolation and treatment. Yet fears about substandard care and a lack of confidence in health workers deterred symptomatic patients in Sierra Leone from visiting clinics. Instead, individuals hid sick family members and evaded testing and contact-tracing efforts (Abramowitz et al. 2016, 24).

To test whether our interventions contribute to the health system's resilience, we ask whether they affect reporting of Ebola cases. We use a deidentified database maintained by the GoSL and CDC to construct weekly counts of tested and confirmed patients in small administrative units called sections. We focus on the 160 sections that contain a single clinic from the experimental sample, which permits unambiguous coding of each section's treatment status. Pooling the two interventions, we estimate that they substantially increased reporting, by 62%. Although we cannot reject the null hypothesis that the interventions have statistically equivalent effects, qualitatively, we see a larger increase in reported cases in sections containing clinics under CM.

We attribute increased case counts to reporting behavior, not Ebola transmission. The programs increase all types of cases—confirmed cases, and cases where patients test negative for the virus. We also rule out nosocomial transmission (i.e., exposure to infected patients in a clinical setting) in 99% of cases, based on the timing of symptom onset and reporting. We thus interpret the increase in reported cases (including confirmed cases) as a critical

step toward containment: a back-of-the-envelope calculation (per Pronyk et al. 2016) suggests that this increased reporting reduces the virus's reproduction rate (R_0) by around 19%.

These findings align with our results prior to the crisis: improvements in the perceived quality of care encourage reporting during the epidemic. In particular, we show that general utilization, satisfaction with public health workers, and confidence in the effectiveness of Western ("white-man") medicine relative to traditional healers all significantly increase in treatment areas in the 160 clinics used in our Ebola analysis. We combine these measures into a perceived quality of care index. Instrumenting that index with our randomized treatment assignment (per Kling, Liebman, and Katz 2007), we find that a one standard deviation change in the perceived quality of care increases Ebola reporting by 0.39 cases per section-week (under the strong assumption that the effects of the treatments only operate through this channel). Separating the two interventions, we find that CM has larger (first-stage) effects on our perceived quality of care index, consistent with its larger effects on reporting noted above.

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

Beyond reporting, we examine mortality among Ebola patients. In sections with CM clinics, we observe a decline in mortality: 1 patient dies for every 10 who report, compared with 1 in 4 in sections with control clinics. This result is conditional on reported cases. Thus our estimates again suggest that CM generates benefits through a channel beyond utilization (e.g., through changes in the quality of administered care). Because improvements in health outcomes are concentrated in CM clinics, both under normal and crisis conditions, the direct community involvement under CM may spur a larger and sustained change in providers' behavior and the resulting quality of health services.

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

A large body of literature addresses the challenges of motivating frontline bureaucrats responsible for delivering services. Community monitoring has been employed across a variety of sectors, including education (Banerjee et al. 2010; Pradhan et al. 2011; Barr et al. 2012; Andrabi et al. 2018), corruption (Fiala and Premand 2018; Olken 2007), and health. In the health sector, CM appears to have larger effects in contexts with poor baseline health outcomes and services, such as in Uganda in 2005 (Björkman and Svensson 2009), India (Mohanan et al. 2020), and our study in Sierra Leone. Encouragingly, Björkman Nyqvist, de Walque, and Svensson (2017) find that these effects persist: following up on Björkman and Svensson (2009), they find lasting treatment effects on health outcomes over the longer run. By contrast, community monitoring may not work as well when baseline health conditions are better (for a recent study in Uganda, see Raffler, Posner, and Parkerson 2019).

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

On the demand side, a growing body of work finds that trust affects the utilization of public services, particularly health care. Fear and distrust deter patients from using health systems over

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

a long horizon (Alsan and Wanamaker 2018; Lowes and Montero 2018). Patients' trust may be particularly important amid public health crises, when they face choices about whether to voluntarily report for medical testing or honor a quarantine. Our findings reinforce work in Liberia (Blair, Morse, and Tsai 2017; Morse et al. 2016; Tsai, Morse, and Blair 2019) and the Democratic Republic of Congo (Vinck et al. 2019), which finds that trust in government affected clinic utilization during those countries' Ebola crises.[3] This research is also echoed in commentary on the COVID-19 pandemic, with experts arguing that distrust of public health officials undermines containment efforts.[4]

The rest of our article is structured as follows. Section II describes the study context, experimental design, and details of the two interventions. Section III introduces our sampling procedure, the survey and Ebola case data, randomization, and empirical strategy. Section IV presents our findings under normal conditions and the longer-run effects under the Ebola crisis and also discusses cost-effectiveness. The final section concludes. The appendix figures and tables that we reference are included in an Online Appendix.

II. Health Care in Sierra Leone

II.A. 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. Online Appendix Figure A.1 displays Sierra Leone's per capita health expenditure and under-five mortality in 2010 relative to other countries that the World Bank classified as low income. Located in the upper-right quadrant, the country spent more and performed worse than countries at a comparable level of economic development. Western-style health care is provided primarily

3. Our work also 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 and Rasul 2019), and how government responds differently when faced with these types of crises (Maffioli 2018).

4. Wen (2020, https://www.washingtonpost.com/opinions/2020/01/22/governments-need-peoples-trust-stop-an-outbreak-where-does-that-leave-us) 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."

through government-run clinics and hospitals; private and NGO-sponsored facilities are scant (Denney and Mallett 2014). Government facilities operate alongside traditional village birth attendants and healers. Our study focuses on primary health clinics—the first points of contact for patients in towns and villages—that each serve populations of 500 to 10,000 (UNICEF 2014). These clinics typically focus on maternal and child care, providing services such as antenatal care, supervised deliveries, postnatal care, family planning, growth monitoring for under-five children, and immunization. In addition, there is some focus on health education and management of minor ailments, as well as referral of more serious medical conditions to larger facilities (MOH 2017).

In an effort to reduce child and maternal mortality, the GoSL launched a free health care initiative in 2010, removing fees for pregnant and lactating women and children under the age of five. The policy simultaneously increased pay for government health care workers; at the time, 30%–50% 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.

Primary health clinics continued to operate during the Ebola crisis: a UNICEF (2014) facility survey in October 2014 (four months after the first confirmed case in Sierra Leone) found that only 4% of clinics were closed. In addition, clinics remained largely accessible: the GoSL implemented short lockdowns, most prominently a three-day nationwide quarantine between September 19 and 21, 2014, that banned all travel. We know of no additional travel bans within our study area that affected clinic access. Levy et al. (2015, 753) report that "early assessments [from October 2014] found that many [Ebola] patients were first seeking care at local [clinics]." Concerned that these clinics lacked the training and equipment to properly isolate and care for Ebola patients, clinic staff were rapidly trained on infection prevention and control and outfitted with personal protective equipment. By early December 2014, 81% of health care workers in Sierra Leone had received training (see Online Appendix Table E.1); by late December 2014, training had reached 98% (Levy et al. 2015). Case studies suggest that clinic staff and community health workers were providing "no-touch" treatment for dehydration and fever and engaged in social mobilization and disease surveillance (Vandi et al. 2017). While training and the

disbursement of protective equipment filled important knowledge and resource gaps, UNICEF's (2014) survey found that 90% of clinic staff felt that fear and misconceptions were "the main challenge confronted by the health system in fighting Ebola."

II.B. 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 frontline health care workers. Without incentives tied to service delivery, the government worried that nurses would miss work or continue to charge illegal fees or inflated drug prices—barriers to service provision that the free health care initiative intended to eliminate.

With World Bank support, the GoSL contracted with three international NGOs to implement two interventions in 170 clinics across four districts. Plan International worked in Bombali district, Concern Worldwide in Tonkolili district, and the International Rescue Committee worked in Bo and Kenema districts.[5] The four districts bisect Sierra Leone from north to south (see Online Appendix Figure A.2) and cover just over 30% of Sierra Leone's population.

1. Community Monitoring (CM). The CM intervention was modelled on Björkman and Svensson's (2009) "Power to the People" approach in Uganda in 2005. The intervention attempts to mobilize "client power," providing patients with information and a forum to demand accountability from frontline staff (World Bank 2003). It convenes users and providers to discuss problems around local health service delivery and agree on 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-five mortality, vaccination rate, percentage of births in a health facility, and completion of four antenatal visits). These were constructed from administrative

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

data provided by the Ministry of Health and Sanitation and compared with 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 inpatient deliveries, and parents failed to complete the vaccination courses for children.

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

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

Finally, facilitators held follow-up meetings three, five, and nine months after the initial interface meeting to revisit the action plan and monitor progress. These meetings were held jointly with the community and clinic staff, and each side rated the extent to which the other side had made progress on their

commitments. The research team monitored almost all CM clinics at some stage of the intervention.[6]

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

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

To encourage truthful reporting of indicators, clinics across all treatment groups were informed that their patient registers would be audited at baseline and endline, and clinics with fraudulent entries would be disqualified from the competitions. Each audit involved randomly selecting 30 patients from the clinic register (corresponding to 15 patients per study community) and visiting each individual to verify their recorded visit date and purpose. None of the audits uncovered ghost patients or manipulated entries in the clinic register.[7]

6. We randomly selected and observed half of the first interface meetings. Meetings typically lasted three to four hours; average meeting attendance ranged from 52 people in Kenema district to 68 in Bombali district and included representatives from the clinic, traditional authorities, and a larger number of community members (with roughly equal representation of men and women). We also monitored the three-month follow-up meeting for nearly all clinics where we did not observe the initial interface meeting. In the three-month follow-up, average meeting attendance was only slightly lower than at the first interface meetings.

7. 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

The implementing NGOs took a number of steps to build broad awareness of the competition. First, they met with district health officials to explain the competition. Second, they advertised the competition extensively in and around the clinics: posters were placed inside the clinic and at high-traffic locations, such as schools and chiefs' homes. Third, the NGOs met with each of the clinics individually to explain the competition and answer relevant questions. During this meeting, they encouraged clinic staff to develop action plans to identify opportunities to improve service provision, without divulging which indicators would be used for the actual rankings. Finally, they held follow-up meetings at the clinics three and six months after the initial meeting to remind staff about the competition. The research team monitored at least one meeting in 49 out of 85 (58%) NFA clinics.

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

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

III. Design and Methods

III.A. Sampling

1. Clinics, Communities, and Households in the Full Experimental Sample. The districts in our study include 318 primary health clinics. We sampled 254, such that all sampled clinics were separated by at least three kilometers to minimize spillovers. As a result, the average distance to the next nearest clinic in our

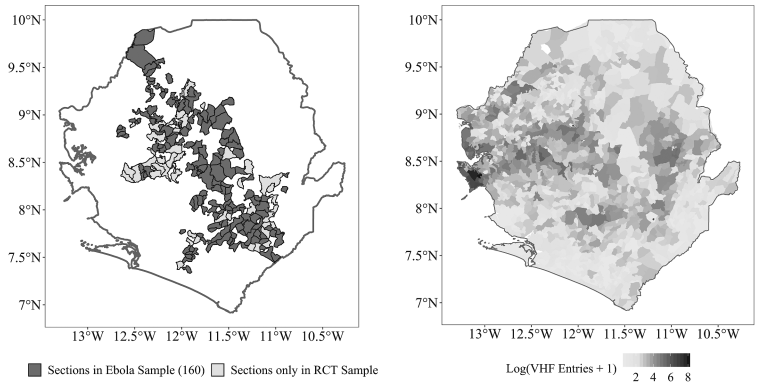
sampled clinic registers from clinics in all study clinics; however, to reduce data collection costs, we only visited patients to verify details for NFA clinics to ensure that awards were handed out correctly. Verification in CM and control clinics would not differentially affect reporting, since the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

sample is 10 kilometers. In Sierra Leone, each clinic has a defined catchment area (a roughly three-kilometer buffer around a clinic) that prioritizes the communities it serves. Individuals are also administratively assigned to a specific clinic, which, combined with high travel costs, discourages the use of more distant clinics. At baseline, the average clinic in our sample had just over two staff members present, reported being open six days a week, and saw roughly 450 patients a month. Over 80% of clinics had walls and roofs in good condition, accessed piped or protected water, and stocks of basic medications (e.g., oral rehydration salts and antibiotics); yet only 10% had functional electrical lighting.

We randomly sampled two communities from each clinic's catchment area. As shown in the CONSORT diagram in Figure I, 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 15 individuals in each community and administered a shorter user feedback survey focused on recent health episodes, service provision, and satisfaction. For the endline, we resurveyed the 5 households that took the baseline household survey. We also randomly selected 5 of the 15 individuals who took the user feedback survey at baseline. This generates a sample of 10 households per community at endline (5,080 households). The households in our sample are poor: at baseline, 74% lived in homes with mud floors and wooden walls, 24% had no toilet facility, another 58% used a pit latrine, only 20% owned a mobile phone, and 62% had no formal education.

2. Blocking and Randomization. We grouped the 254 clinics in our sample into matched triplets using Greevy and Beck's (2016) non-bipartite matching algorithm. Clinics in a triplet fall within the same district and exhibit similar levels of utilization and performance at baseline.[8] Blocking on matched triplet, we randomized 84 clinics into control, 85 into CM, and 85 into NFA.

8. We exactly match clinics by district and clinic type (maternal and child health post or community health post). We then select matches based on the Mahalanobis distance between eight indicators specified by the Ministry of Health and Sanitation: completion of first-year vaccinations, institutional deliveries, completion of fourth antenatal care visit, charging of fees for maternal and under-five services, nurse absenteeism, staff attitude, maternal mortality, and under-five mortality.



(A) Sections Used in Ebola Analysis (B) Heat Map of VHF Cases by Section

FIGURE II

Mapping of Ebola Cases and Sample

Panel 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. Panel B: The number of entries by section in the Viral Hemorrhagic Fever (VHF) database was 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.

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

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

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

III.B. Data Collection

1. Survey Data on Health Clinics, Services, and Outcomes. Baseline surveys were administered in September 2011, and endline surveys in May and June 2013 (see Figure I for a timeline). We rely on three survey instruments: first, surveys at each clinic, in which enumerators audited the staffing, cleanliness, drug stocks, and registers of clinics; second, surveys of leaders in each community regarding amenities, relations with the clinic, and community development; and third, household surveys that captured attitudes, behaviors, and outcomes related to health and economic well-being.

We filed an analysis plan to examine the survey outcomes at the AEA RCT registry.[9] The plan defines 10 outcome families, including subcomponents that make up each family. We flag and explain any subsequent deviations in Online Appendix B.1.

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

9. 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 on at the beginning of the study: institutional delivery, antenatal care visits, immunization, illegal fees, nurse absenteeism, staff attitude, maternal and under-five mortality, utilization, and anthropometric outcomes. We did not examine other outcomes from the household data, or any outcomes from the clinic or community data.

where 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 thus represent standard deviation changes relative to the control group. Following Kling, Liebman, and Katz (2007), in case y_{ik} is missing but another subcomponent of the family is measured, we impute the mean from the same treatment arm and survey wave. Some subcomponents, for example, those that relate to childbirth, are only defined for a fraction of respondents. For that reason, we do not impute values when estimating treatment effects for individual subcomponents. To demonstrate that the imputation is innocuous when looking at effects on families, we follow Kling and Liebman (2004) and Casey, Glennerster, and Miguel (2012) and aggregate treatment effects across the subcomponents of each family using seemingly unrelated regressions (SURs). These results (reported in Online Appendix D) are qualitatively similar across all specifications.

Below we describe each outcome family; Online Appendix B.1 provides additional detail on each family's subcomponents, and Online Appendix Table B.2 includes descriptive statistics of each variable at endline.

i. **General utilization** measures the number of episodes in which individuals seek care at a Western-style clinic, including in response to four of the most common health needs addressed at primary health units—childbirth in the past year, antenatal or postnatal care, vaccination, or any illness or injury, as well as a residual category of any other type of consultation in the past month.[10] While most utilization occurs in response to specific health needs (as regular health check-ups are not common in our setting), the residual category helps generate a comprehensive measure of utilization. Utilization of a Western-style clinic reflects the decision to seek care at a formal clinic, rather than visiting a traditional healer or spiritual leader or forgoing any type of care.[11] The Western-style

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

11. Our analysis plan specified examining utilization of traditional religious healers. However, due to an error in survey design we do not have complete utilization data for traditional healers for all health episode types

clinics utilized by respondents are overwhelmingly government-run clinics; utilization of private or NGO-run clinics constitutes a small share of utilization (3%).

ii. **Maternal utilization** is measured among women who gave birth in the year before the endline survey. The family includes two outcomes: an index of the number of times a woman sought antenatal care (ANC) or postnatal care (PNC), and an indicator for whether the woman gave birth in a Western-style clinic.[12]

iii. **Health outcomes** are measured at the household level. The family includes four measures related to child health: under-five mortality over the past six months; under-five illnesses over the past month (e.g., malaria or diarrhea); under-two vaccine completion; and under-five child wasting, measured using the weight-for-length ratio.[13] The family also includes three other variables: two related to childbirth, maternal mortality over the past 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.

iv. **Satisfaction** is measured at the household level. The family includes three outcomes measured on a four-point Likert scale from "very unsatisfied" to "very satistfied": the respondent's satisfaction with their family's health, satisfaction with public health workers (i.e., clinic staff), and—among households with at least one member utilizing a Western-style clinic in the last year (approximately half of the sample)—satisfaction with the care they

(see note A2 in Online Appendix B.1). We therefore focus on utilization of Western-style clinics in the main results. We do have utilization data for both Western clinics and traditional healers for one specific episode type, namely, illness/injury episodes. We conduct robustness checks using this episode type alone in Online Appendix Tables D.21 and D.22.

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

13. 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.

received.[14] 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.

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

vi. **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.

vii. **Community support** 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.

14. 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.

viii. **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 past six months, contributions to local development projects over the past year, their self-reported ability to address problems collectively over the past year, and turnout for the local and national elections in November 2012.

ix. **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.

x. **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 past month.

2. Ebola Case Data. We rely on a deidentified version of the Epi Info Viral Hemorrhagic Fever (VHF) database, which was the primary data management system used to track the Ebola outbreak in Sierra Leone.[15] The Ministry of Health and Sanitation, with support from the CDC, implemented and maintained the VHF database through the end of the epidemic, and McNamara et al. (2016, 39) describe it as "the most comprehensive epidemiologic and laboratory data on Ebola cases available in Sierra Leone." The VHF compiles patient information, their lab results, and whether they died. Patients could enter the VHF through walk-in visits to health centers, as well as surveillance activities (e.g., contact tracing) (Owada et al. 2016). As noted already, the VHF reflects reported cases, rather than actual Ebola incidence—a particularly important outcome for stopping contagion and containing the epidemic (Enserink 2014).

We use information on patients' residences to geocode cases to sections. (The location of symptom onset is recorded for only a subset of cases; when it is not missing, it matches the patient's

15. The Njala University Ebola Museum and Archive facilitated access to this database.

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

In the 160 sections that constitute our primary Ebola sample, the VHF includes 2,045 case entries, which are classified into four types: 1,623 negative cases where Ebola has been ruled out; 269 confirmed cases; and two residual categories that are never confirmed with lab tests: 134 suspected cases which display Ebola symptoms and/or have had contact with potentially infected individuals or animals, and 19 probable cases, which meet the criteria for a suspected case and were either screened by a clinician or died and have an epidemiological link to a confirmed case.[16] Given our interest in reporting, our main dependent variable is the count of total cases (the sum across the four case types) aggregated to the section-week. We use the date when a case is first entered in the VHF database to determine the week. Online Appendix Table E.3 presents descriptive statistics for total and confirmed cases.

16. Suspected cases include (i) the onset of high fever and contact with a suspected, probable, or confirmed individual or a dead or sick animal; (ii) 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 hiccups; any person with inexplicable bleeding; or any sudden, inexplicable death. Suspected and probable cases may have died prior to a lab sample being collected; alternatively, administrative issues may have led to tests being overlooked or not entered into the VHF.

III.C. Manipulation Checks and Balance

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

Finally, as specified in our analysis plan, we asked survey respondents whether key program activities took place in their communities or clinics. In Online Appendix Table C.1, we find that 86% of leaders in CM communities report "meetings held by IRC, Plan, or Concern to discuss how the clinic and community can work together to improve service delivery in this community." This is roughly double the rate compared with control communities and suggests broad awareness of interface meetings in CM communities. Note, however, the high control mean (44% of control communities also report a community meeting). This likely reflects confusion, as NGOs commonly convene community meetings across rural Sierra Leone. NFA communities also report an increase on this measure (of 12 percentage points), albeit significantly less than CM. Given the monitoring we describe, we do not attribute this to contamination. Rather, the NFA protocol also involved meetings convened by these NGOs to develop action plans to improve service delivery in the community. The leaders answering the community survey in NFA communities may have (understandably) responded affirmatively; some of these leaders actually participated in the meetings convened at NFA clinics.

Online Appendix Table C.2 offers a similar story: over 81% of staff at NFA clinics report participating in a competition, which is five times the rate among control clinics. Yet smaller shares of staff at clinics in both control and, to a greater extent, CM also report competing. (The difference across treatment arms is 46 percentage points, which is substantial and statistically significant.) We visited nearly all CM clinics and found no NFA programming. Thus we attribute these responses to misinterpretation:

staff in CM clinics may have interpreted the scorecards (and the comparison to other clinics in the same district) as inviting competition across facilities.

2. Balance in the Full Experimental Sample. Online Appendix Table C.3 reports balance across prespecified covariates. Most variables are individually balanced across treatment arms. We find that the number of injuries or illnesses reported is lower in both CM and NFA relative to control; and in CM, household size is slightly smaller, there is lower trust of village health committees (VHCs), and fewer households report a recent childbirth. However, if anything, we expect that such imbalances make it harder to find effects on general and maternal utilization. We also find that NFA communities have better cellphone coverage, and individuals are less likely to belong to the Temne ethnic group, less likely to believe what a doctor told them, and have a higher level of educational attainment.[17] Given two treatment arms and a control group, we use a multinomial logit model to assess whether the baseline covariates jointly predict treatment assignment. At the bottom of the balance table, we report p-values from chi-squared tests of joint orthogonality, following Özler et al. (2018). These tests suggest that the covariates together are not jointly significant. We also report results where we control for baseline imbalance in Online Appendix Table D.24.

3. Balance in the Ebola Sample. Online Appendix Table E.4 reports balance checks for the 160 sections in our Ebola sample. Some of the imbalance observed in the full sample carries over to this subset, while a small number of variables are imbalanced in the subset but not in the main sample (i.e., CM and NFA communities are less likely to have a prominent village member in the household, while CM communities are more likely to have a motorable road, lower trust, and more members of the Temne versus Mende ethnic group). But again, chi-squared tests of joint

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

orthogonality presented at the bottom of this table indicate that the covariates are not jointly significant in the Ebola subsample. We also report results where we control for baseline imbalance when examining Ebola outcomes in Online Appendix Table E.21.

III.D. Specifications

1. Survey Outcomes. Our main specification is the following ANCOVA-type model:

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

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. If this variable is missing for a given village, we incorporate imputed values and a separate indicator variable for these observations, which controls directly for the imputation effect while enabling us to retain these observations in our sample. Thus we estimate missing indicator ANCOVA models. We use the village average because our baseline survey included a smaller sample of households due to cost considerations. Moreover, some outcomes are only defined for individuals who recently experienced a given health episode. Some households surveyed at endline that experienced relatively infrequent health episodes, such as childbirth, would likely not have also experienced the same episode at baseline. For both reasons, controlling for a household's baseline outcome would reduce the size and representativeness of our sample, and we therefore use the village-level average. When y is a subcomponent of an outcome family, we use the family-level outcome to compute the baseline average.[18] We cluster our standard errors on clinic, the unit of randomization. We also estimate a variant of equation (1) in which we combine the CM and NFA treatments into one pooled treatment indicator.

18. This decision is motivated by two features of our data: first, some subcomponents are only measured at endline; second, for some subcomponents 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 pretreatment covariate, we include the average family-level outcome. This represents a slight deviation from the analysis plan but does not affect any of our conclusions.

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 (i) across outcome families and (ii) across subcomponents in each family.[19]

2. Ebola Outcomes. We assess the impact of the CM and NFA interventions on reported cases for the 160 sections in the Ebola sample (described in Section III.A). We observe counts of reported cases in each section in every week from August 10, 2014, to October 18, 2015. We restrict attention to the period from September 2014 through April 2015, when Ebola transmission was a real threat in our study area; only three confirmed cases were reported during May and October 2015.[20]

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.[21]

19. In the analysis plan, we specified controlling for the FDR only across some families (denoted "primary families") and, within those families, only across some subcomponents. 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 subcomponents.

20. In Online Appendix Table E.5 we extend the panel back to August 2014 and replicate our primary results from Table III.

21. 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.

We amend equation (3) to detect spillovers in 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).

IV. Results

IV.A. Effects Prior to the Ebola Crisis

Our tables follow a common format. Column (1) provides the control mean and standard deviation at endline, which by construction are zero and one exactly when looking at family-level mean-effects indices. Column (2) presents the average treatment effect (in standard deviation units) when pooling the treatment arms. Columns (3) and (4) separately estimate the average treatment effects for CM and NFA, respectively. Column (5) shows the difference between the average treatment effects in CM and NFA. Column (6) provides the F-test for the joint null hypothesis of no effect from either treatment. Finally, column (7) gives the sample size used for each regression. The tables in Online Appendix D.1 and D.2 aggregate across the subcomponents of families using SUR.[22] These estimates, presented in the first row of every table, show qualitatively similar results to the mean-effect indices. The remaining rows in these appendix tables also show treatment effects on individual subcomponents.

Table I examines the interventions' effects on utilization, satisfaction, and health outcomes. Prior to the Ebola outbreak, both programs increase general utilization: the pooled treatment effect is 0.11 standard deviations, with statistically indistinguishable effects across the different two arms. Individuals in the control group used a Western-style clinic for roughly 1 (0.96) health episode (see Online Appendix Table D.1); the treatments increase utilization of such facilities by about 5%. When we focus attention on the utilization of government-run clinics in Online Appendix Table D.20, our effects increase to 7.4% and 6.0% for CM and

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

TABLE I

UTILIZATION, SATISFACTION, AND HEALTH OUTCOMES

	Control mean (1)	Pooled (2)	CM (3)	NFA (4)	Difference (5)	Joint F-test (p) (6)	N (7)
General utilization	0.000	0.112	0.126	0.099	0.026	7.054	4,496
	(1.000)	(0.031)***	(0.034)***	(0.037)***	(0.033)	(.001)***	
		[0.005]***	[0.003]***	[0.032]**			
Maternal utilization	0.000	0.061	0.175	– 0.043	0.218	4.128	888
	(1.000)	(0.064)	(0.077)**	(0.076)	(0.081)***	(.017)**	
		[0.327]	[0.068]*	[0.548]			
Satisfaction	0.000	0.097	0.086	0.108	– 0.022	2.840	5,052
	(1.000)	(0.041)**	(0.047)*	(0.048)**	(0.048)	(.060)*	
		[0.041]**	[0.108]*	[0.049]**			
Health outcomes	0.000	0.064	0.166	– 0.039	0.205	8.105	5,053
	(1.000)	(0.051)	(0.055)***	(0.060)	(0.055)***	(.000)***	
		[0.265]	[0.014]**	[0.548]			

Notes: Treatment effects are estimated using missing-indicator ANCOVA, controlling for the community-level average of the outcome family index at baseline and matching-triplet fixed effects. Column (1) reports standard deviations in parentheses. Columns (2)–(5) report robust standard errors, clustered by clinic. Multiple-inference corrected q-values that adjust for the false discovery rate within treatment arm across all 10 prespecified 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. * significant at the 10% level, ** significant at the 5% level, *** significant at the 1% level.

NFA, verifying that the interventions boosted utilization of the targeted clinics.[23] We are only able to compare the utilization of Western-style clinics and traditional healers for one type of health episode, illness and injuries. In Online Appendix Tables D.21 and D.22, we detect a shift (of roughly equal magnitude) away from traditional or religious care and toward Western-style clinics.

We interpret utilization as a revealed-preference measure. Our results thus suggest that individuals responded to improvements in the perceived quality of care offered at CM and NFA clinics.[24] Perceived quality of care encompasses changes in actual quality, as well as beliefs about the care provided at clinics. Increased utilization (and the improvements in satisfaction we describe below) are consistent with improvements along one or both dimensions.

The CM arm also shows additional effects on maternal utilization: among women who gave birth in the year before the endline survey, maternal utilization increases by 0.13 standard deviations. There is no equivalent effect for NFA. Online Appendix Table D.2 shows that the increase in maternal utilization is driven by more deliveries at Western-style clinics: the probability of giving birth in these facilities is 0.83 in control areas; CM boosts this rate by 9 percentage points (11%). We estimate no effect on antenatal and postnatal visits.

Consistent with the perceived quality of care improving, the third row of Table I shows increased patient satisfaction. We find similar effects in both arms. Pooling the treatments, satisfaction increases by about 0.10 standard deviations, largely driven by increases in respondents' satisfaction with their own health and the performance of health workers (see Online Appendix Table D.3).[25] We generally see high baseline levels of satisfaction,

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

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

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

though 17.5% of respondents report they are somewhat or very unsatisfied with public health workers. Unsurprisingly, the programs' effects on this outcome reflect improvements among households with low baseline levels: splitting our sample into thirds using baseline responses, we find treatment effects on satisfaction with public health workers only in the bottom two terciles (results available on request). All households are asked about their satisfaction with public health workers; thus these effects can arise from improved experiences at clinics, as well as hearing neighbors' positive assessments.

The quality of care administered at clinics (e.g., the time that nurses spend on diagnoses or treatment plans) may also improve. Unfortunately, patient–provider interactions are difficult to measure (for an exception, see Das et al. 2016). We look instead at child health outcomes, assuming that these respond to the actual quality of health care, not just parents' beliefs about the quality of clinics. The fourth row of Table I shows that CM leads to an improvement in health outcomes (0.17 standard deviations). This is driven by significant improvements in child health. As shown in Online Appendix Table D.4, the likelihood that a child under five dies in CM falls by 0.015 relative to the control mean of 0.039, a 38% effect. In addition, child weight-for-length increases by 0.16 z-score units and is significant at the 10% level, though this individual indicator loses significance after FDR adjustments. It is worth noting that the magnitudes of these effects are sizable and qualitatively similar to those uncovered by Björkman and Svensson's (2009) evaluation of community monitoring in rural Uganda. Finally, the effect size for vaccine completion is also large, corresponding to a 10% increase, though the change is not statistically significant. Improvements in health outcomes could reflect increased utilization as individuals access more treatment and preventive care at clinics. However, general utilization increases in both arms, while health outcomes only improve under CM. In addition, the improvements in maternal utilization under CM reflect an increase in institutional deliveries over the last year rather than increases in ante- or postnatal visits (see Online Appendix Table D.2). Decisions by recent mothers (of which there are only 888) to deliver in clinics are unlikely to affect under-five mortality among the much larger set of households

across CM and NFA suggest that social desirability bias is unlikely to drive the estimated effects.

included in our survey. These patterns suggest that an additional channel, i.e., greater effectiveness of health services, contributes to improved health outcomes under community monitoring.

We next study effects on other families that could influence clinic utilization and health outcomes. In the top panel of Table II, we examine the quantity of health services and community contributions to clinics, because either a larger menu of services or groundswell of support could draw in patients and improve their outcomes. The first row of Table II shows no significant effects on health service delivery.[26] However, in Online Appendix Table D.5, we find divergent results for the indicators that measure the quantity of services versus those reflecting their quality. We find no effects on the availability of drugs, medicines in stock, or staff presence.[27] Yet we see a 45% reduction in unpleasant staff behavior in NFA areas, although the effect is not statistically significant. The coefficient for staff attitude also suggests improvements, though it loses significance after the FDR adjustments.[28] These effects, which point to more positive patient–provider interactions, suggest that the effectiveness of health services may also have improved with NFA, though the effects are not strong and do not suffice to improve health outcomes.

When we examine effects on clinic organization and services in the second row of Table II, we do not see any significant

26. We observe null effects on health service delivery despite including the "satisfaction with care" and "would return to clinic" variables, which are also subcomponents 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 on request.

27. We observe a positive effect of NFA on absenteeism in Online Appendix Table D.5. This is likely an artifact 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 posttreatment bias pushes toward 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. Online Appendix Table D.23 shows precise null effects on this outcome.

28. There is a small (1.5%) and marginally significant increase in whether people would return to the clinic under CM. Note that ceiling effects may limit our ability to detect improvements using this measure: nearly all (97%) patients in control areas report that they would return.

TABLE II

SUPPLY-SIDE MEASURES AND COMMUNITY SUPPORT

	Control mean (1)	Pooled (2)	CM (3)	NFA (4)	Difference (5)	Joint F-test (p) (6)	N (7)
Supply-side measures and community support							
Health service delivery	0.000	0.026	0.041	0.022	0.019	0.141	2,877
	(1.000)	(0.057)	(0.079)	(0.060)	(0.071)	(.869)	
		[0.537]	[0.504]	[0.619]			
Clinic organization and services	0.000	0.104	– 0.004	0.213	– 0.216	0.929	254
	(1.000)	(0.149)	(0.175)	(0.176)	(0.184)	(.397)	
		[0.437]	[0.649]	[0.237]			
Community support	0.000	0.033	0.044	0.021	0.023	0.079	508
	(1.000)	(0.095)	(0.112)	(0.109)	(0.113)	(.924)	
		[0.537]	[0.504]	[0.619]			
Community development projects and infrastructure							
CDPE	0.000	0.231	0.202	0.261	– 0.059	3.849	508
	(1.000)	(0.085)***	(0.102)**	(0.101)**	(0.110)	(.023)**	
		[0.034]**	[0.095]*	[0.032]**			
Water and sanitation	0.000	0.156	0.093	0.219	– 0.126	4.566	5,053
	(1.000)	(0.063)**	(0.073)	(0.073)***	(0.072)*	(.011)**	
		[0.041]**	[0.202]	[0.030]**			

Notes. Treatment effects are estimated using missing-indicator ANCOVA, controlling for the community-level average of the outcome family index at baseline and matching-triplet fixed effects. Column (1) reports standard deviations in parentheses. Columns (2)–(5) report robust standard errors, clustered by clinic. Multiple-inference corrected q-values that adjust for the false discovery rate within treatment arm across all 10 prespecified 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. * significant at the 10% level, ** significant at the 5% level, *** significant at the 1% level.

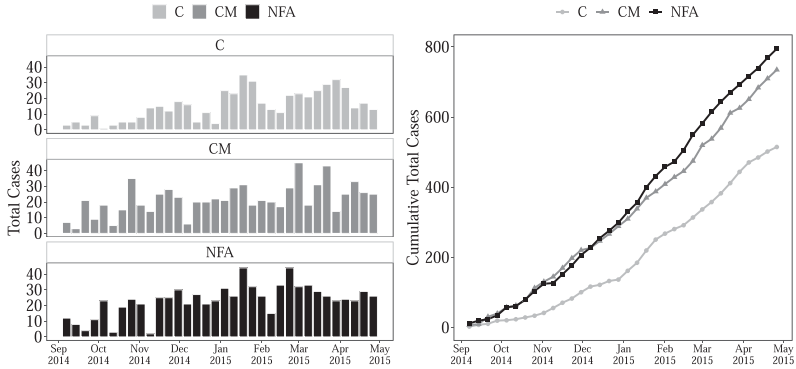
effects at the family level or for individual subcomponents (see Online Appendix Table D.6). This is not surprising because neither intervention provided additional resources to clinics; the government's interest in the evaluation was understanding how to extract more effort from health workers under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

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

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

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

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



(A) Weekly Counts (B) Cumulative Counts

FIGURE III

Total Ebola Cases by Treatment

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

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

IV.B. Longer-Run Effects during the Ebola Crisis

Roughly one year after our endline survey, the first confirmed Ebola case was recorded in Sierra Leone. We turn to examining the longer-run effects that the interventions had during the epidemic, including reporting of Ebola cases, as well as mortality among Ebola patients.

1. Effects on Reporting. The treatment effects on reported Ebola cases are apparent in Figure III: the left panel presents the sum of total reported cases in each week by treatment arm;

TABLE III
REPORTED EBOLA CASES PER SECTION PER WEEK

	Control mean (1)	Pooled (2)	CM (3)	NFA (4)	Difference (5)	N (6)
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 are estimated using OLS including matching-triplet and week fixed effects. Column (1) reports the standard deviation in parentheses. Columns (2)–(4) report robust standard errors, clustered by section, in parentheses. The difference column reports the difference between the CM and NFA coefficients; the standard error is computed using the delta method. The bottom panel employs the inverse hyperbolic sine transformation, $\text{IHS}(y) = \log(y + \sqrt{1 + y^2})$. Total cases include confirmed, negative, suspected, and probable cases. * significant at the 10% level, ** significant at the 5% level, *** significant at the 1% level.

the right panel is the cumulative count of cases during our study period. Between September 2014 and May 2015, we count 515 total cases in control sections; yet in sections with clinics receiving the CM and NFA interventions, 735 and 795 cases are reported, respectively. This difference is even more striking for confirmed cases: only 21 confirmed cases are reported in control sections, whereas 248 are reported in the treated sections (see Online Appendix Figure E.1).[29]

We present regression results using equation (3) in Table III. In the top panel, the outcomes are the raw counts of total, confirmed, and negative cases. The pooled effect implies a 62% increase in the average number of total cases. The effect is smaller and less precisely estimated for NFA ($p = .13$), which is consistent with our pre-Ebola findings, where we observe more limited effects of NFA on utilization outcomes, especially in the Ebola

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

sample (see second row of Table V). Nonetheless, we cannot reject the null hypothesis that CM and NFA have equivalent effects. In the bottom panel of Table III, we find similar effects using the inverse hyperbolic sine (IHS) transformation of counts. The coefficient on the pooled treatment implies a 40.3% increase (see Bellemare and Wichman 2020); the effect of NFA on confirmed cases becomes significant in this specification. We interpret the increase in Ebola cases as reflecting reporting behavior by individuals, rather than effects on transmission: in treatment areas, more individuals reported into clinics to get tested and, if needed, get treatment.

To both improve patient survival and contain the epidemic, it is particularly important that infected patients report. We find large increases in the average number of confirmed cases reporting in treated sections: for every confirmed case in control, we count five confirmed cases in treated sections (based on the pooled treatment effect in the second row of Table III). We consider the implications of these estimates for the spread of the epidemic. Back-of-the-envelope calculations, following the method employed by Pronyk et al. (2016), suggests that increased reporting by infected individuals reduced the disease's reproduction rate (R_0) by 19% (see Online Appendix E.10).

We also observe increases in cases that test negative for the virus. This reinforces our claim (which we substantiate further below) that the effects on total cases reflect increased reporting by individuals seeking testing and treatment.[30]

2. Effects on Patient Deaths. We posit that individuals report more in sections with treated clinics due to improvements in the perceived quality of care. As with the pre-Ebola period, we lack objective measures of the quality of care administered at clinics during the epidemic. However, we are again able to look at a health outcome, which will partly be shaped by the quality of care that was administered—namely, Ebola patient deaths. Sierra Leone lacks vital statistics data, so we can only analyze mortality for cases in the VHF database. We regress the number of deaths in each section-week on the total number of cases reported in the current and previous week and the interaction of that caseload with treatment. We opt for the caseload over the current and

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

TABLE IV
EFFECT ON PATIENT DEATHS

	Dependent variable:	
	Patient deaths	
	(1)	(2)
Total cases in last two weeks	0.245	0.247
	(0.021)***	(0.021)***
Pooled	0.063	
	(0.032)**	
Total cases in last two weeks × pooled	− 0.098	
	(0.043)**	
CM		0.116
		(0.037)***
Total cases in last two weeks × CM		− 0.149
		(0.046)***
NFA		− 0.007
		(0.025)
Total cases in last two weeks × NFA		− 0.019
		(0.032)
Control mean	0.149	0.149
	(0.49)	(0.49)
Observations	5,280	5,280

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

previous week, as Ebola deaths typically occur 6 to 16 days after symptom onset.

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

These conditional-on-positives estimates will be confounded if treatment changes the composition of patients (e.g., their

comorbidities). The increased number of confirmed patients in treated sections should, if anything, attenuate these results. Despite more infected cases reporting, our findings suggest that patients in CM sections enjoyed higher survival rates. One may worry that patients in control simply waited longer to report and, thus, presented with greater illness severity and higher risk of mortality. Yet we show in Online Appendix Table E.7 that treatment does not reduce the number of days between symptom onset and reporting.

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

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

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

First, increased transmission could arise from greater interaction among community members in treatment areas. However, meetings associated with the interventions concluded five months before the first Ebola case in Sierra Leone. Had they continued, meetings are unlikely sites of transmission: Ebola is not an airborne pathogen; it requires direct contact with bodily

fluids (e.g., blood, feces, saliva, vomit).[31] Those facts notwithstanding, the treatments may still have increased interactions outside the home and thus enabled transmission. However, data from contact-tracing efforts do not support this possibility. For a subset of infected patients, caseworkers identify people who may have come into contact with the patient. Through this process, they record how contacts are related to the patient (e.g., neighbor, tenant, brother, grandmother). In the last two columns of Online Appendix Table E.11, we find that contacts outside the nuclear family were, if anything, lower in CM and NFA areas compared with control areas. This pattern is inconsistent with greater interactions and contact outside the family among infected patients in treatment areas.

Second, by increasing the number of individuals reporting into clinics, the treatments could have increased contact between infected and susceptible individuals at these facilities, raising the risk of nosocomial transmission (for an account that relates nosocomial transmission to distrust, see Lowes and Montero 2018). To address this possibility, we compare the dates of symptom onset, reporting, and lab testing. Two features of the Ebola virus are important to note: first, Ebola incubates for 2 to 21 days (8–10 on average) before showing symptoms; second, an individual can only test positive after displaying symptoms. Consequently, symptom onset or positive lab results in the first two days after a patient reports cannot reflect infections due to exposure after the patient reports into clinics. However, for 92% of confirmed cases in our sample, symptom onset occurs prior to reporting, and in 99% of cases (all but two cases), either symptom onset or lab testing occurs within two days of reporting. This indicates that nearly all confirmed cases we count do not result from infections that occur after the case was reported. (The proportions are nearly identical among patients who test negative for Ebola: 89.8% have symptom onset prior to reporting, and 99.4% have onset or lab testing within two days of reporting.)

As further evidence against nosocomial transmission in our sample, Fang et al. (2016) report that infections among health care workers fell precipitously by September 2014 (the start of our Ebola study period), indicating improved awareness and infection control. We continue to find treatment effects in the

31. This is why Glynn et al. (2018) estimate a secondary attack rate of only 18% among individuals living in the same household as a confirmed Ebola patient.

months after a nationwide effort during November and December 2014 to train health care workers in isolation and no-touch treatment (see Online Appendix Table E.1 and E.8).

Third, we conduct a placebo test in which we substitute the nearest out-of-sample neighbor for each section. We find no significant effects (see Online Appendix Table E.9), alleviating concerns that our treated sections are spatially clustered in areas where reporting is higher for reasons unrelated to treatment, such as greater transmission of Ebola.[32]

Finally, we look at the ratio of confirmed to total cases across treatment and control areas to determine whether the interventions increased the share of infected patients among total cases. This ratio is undefined when no cases are reported in a section-week. We therefore take a bounding approach, imputing either all ones or all zeros to observations where the ratio is undefined. Imputing all ones assumes that if cases had been reported, they would have all been confirmed; imputing all zeros assumes that if cases had been reported, none would have tested positive. Online Appendix Figure E.5(a) plots the average ratio of confirmed to total cases across control and treated sections. Looking at either bound, there is no meaningful difference in these ratios, and the confidence intervals overlap throughout the study period.[33]

In Online Appendix E.14, we write down a model to clarify what must be assumed for our results to reflect a change in transmission (as opposed to reporting). For confirmed cases to increase while the share of confirmed to total 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,

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

33. 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. Online Appendix Figure E.5(b) repeats the bounding exercise but uses the ratio of confirmed to confirmed plus negative cases. This exercise delivers the same conclusion, as the number of probable and suspected cases are small and unaffected by treatment (see Online Appendix E.20).

qualitative accounts suggest the crisis deterred unexposed individuals from visiting clinics, even when they had other health care needs (Elston et al. 2016).

4. Addressing Increased Surveillance. Treatment clinics were more exposed to three international NGOs and may have had more communication with the government Ministry of Health. This might have increased "top-down" disease surveillance and thus increased reported cases.[34]

We first examine rates of contact tracing, which is central to disease surveillance efforts. In our control sections, 59% of confirmed cases were subject to contact tracing, compared to just 22% in CM and 24% in NFA (Online Appendix Table E.11). Second, we examine three measures derived from the VHF data which also proxy for top-down surveillance efforts: (i) the probability that a case received laboratory testing to confirm or rule out an infection; (ii) the average number of days that passed between a case being reported and lab testing; and (iii) the number of unique case workers (logged) that entered information into the VHF. In Online Appendix Table E.12, we find no significant differences for these variables across treatment and control.

Next, using data from Sierra Leone's National Ebola Response Center (NERC) and the UN Mission for Ebola Emergency Response (UNMEER), we count the number of Ebola-specific treatment facilities in each section (see Online Appendix E.16). There were three types of specialized facilities: Ebola Treatment Units (ETUs), Ebola Holding Centers (EHCs), and Community Care Centers (CCCs). Only one ETU falls within our sample, and it is located in a control section; Online Appendix Table E.13 shows no significant difference in the counts—either combined or separate—of EHCs or CCCs.[35] Our results in Table III are also robust to dropping the small number of sections that contain

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

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

one or more of these specialized facilities (results available on request).[36]

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 Online Appendix Table E.1). Unfortunately, only aggregate data are available on the roll-out of training, so we cannot date when individual clinics were covered, but given the pace of the roll-out, all clinics in our sample were likely to have received the training around the same time.

5. Additional Checks of Ebola Results. We demonstrate robustness to a number of alternative specifications. In Online Appendix Table E.16, we present estimates using a linear probability model, a Poisson count model, a rare-events logit model, and logged counts (adding 1 to avoid dropping section-weeks with no cases). In the $\log(y + 1)$ transformation, the coefficient on the pooled treatment implies a 41% increase, similar in magnitude to the implied effect of 40.3% in the IHS specification of Table III. 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 .104.[37] In Online Appendix Table E.18, we also estimate a dose-response model that extends the sample to 205 sections, including sections with multiple study clinics, using the proportion of clinics in a section that were treated as the right-hand side variable. We find similar effects under this approach.

To ensure that our results are not driven by a particular place or period, we conduct subsample analysis. We reestimate the pooled effect dropping one matched triplet at a time (Online Appendix Figure E.2), dropping each possible pair of matched

36. 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 Online Appendix Table E.14 we find that treated sections are not significantly further from ETUs, EHCs, or CCCs in the NERC data; the distance from NFA sections to the nearest CCCs is shorter when we use the UNMEER data.

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

triplets (Online Appendix Figure E.3), or dropping each week (Online Appendix Figure E.4). Second, we estimate the effects by month to assess whether our results are driven by a particular moment in the crisis. The pooled coefficient is positive in every month, and we find significant effects in October, December, February, and April (Online Appendix Table E.8). (We find large and significant effects for CM in October 2014 and April 2015; for NFA, in October and December 2014.) The spread of these effects across our study period verifies that our estimates are not driven by any particular period. Probable and suspected cases (which constitute 1% and 6.5% of total cases, respectively) are included in our count of total reported cases. However, these case types often do not involve reporting by individuals; their ambiguous status reflects the absence of a definitive lab test (e.g., confirmed or negative). These cases include, for example, deceased individuals with Ebola symptoms. We separately analyze these cases in Online Appendix Table E.19 and find insignificant and negligible treatment effects.[38] In Online Appendix Table E.20 we also subtract probable and suspected cases from total cases and find similar effects. These checks indicate that estimates in Table III are driven by increases in the number of patients who report and receive testing.

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

We next look for evidence of spillovers between treated and control sections, particularly indications that patients traveled from control to treated sections—reallocation which would amplify our treatment effects on reported cases. Assuming that patients minimize travel costs, spillovers should be largest in treated sections that border (populous) control sections. In Online Appendix Table E.22, we interact our treatment indicator, first, with the number of bordering control sections and, second, with the population (based on 2004 census data) in bordering control sections. If patients from control sections report in adjacent treatment areas, the coefficients on these interactions will be positive; yet our estimates are negative and insignificant. We

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

conduct a number of additional analyses to look for spillovers based on geographic or cultural proximity between treatment and control sections: we look at the distance between clinics (see Online Appendix Table E.23), connections via road networks (see Online Appendix Table E.24) and cultural similarity (see Online Appendix Table E.25). Across these specifications, the point estimates on the interaction terms are all negative, and as a result, the coefficients on the pooled treatment are larger after taking spillovers into account. In addition, imprecision in estimating spillover effects is not a likely source of error in quantifying these spillovers: even when we assume that the true spillover effect is at the "conservative" boundary of its confidence interval and adjust the treatment effect estimates accordingly, they are still large and positive in all cases except one (see Online Appendix E.22).

Finally, confirmed Ebola cases are a relatively rare event, which raises concerns about power. However, note that the standard error on our main result, the pooled treatment effect on total Ebola cases, is 0.083 (Table III), implying that we would have rejected the null hypothesis of no effect at the 95% confidence level for coefficients larger than $0.083 \times 1.96 = 0.163$. Thus, we were powered to detect effects of a magnitude that can reasonably be expected in field studies.

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

Concerns about substandard care are believed to have deterred patients from utilizing clinics during the Ebola crisis. Fearful that seeking care would condemn their loved ones to death, households "engaged in practices of hiding sick family members, running away from local communities, or attempting to manage the course of Ebola within local households and communities" (Abramowitz et al. 2016). If the CM and NFA interventions generated persistent improvements in the perceived quality of health care, this would help explain increased reporting in treated sections.[39] Using our endline surveys but restricting attention to the 160 clinics in the Ebola sample (see Online Appendix

39. 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 make up over 70% of the confirmed Ebola cases.

Table D.26 for estimates using our full sample), in Table V we estimate treatment effects on general utilization, satisfaction with public health workers, and households' beliefs about the effectiveness of Western-style medicine relative to traditional or religious remedies, the primary alternatives to government-run clinics in rural Sierra Leone. (See Online Appendix Table E.26 for effects on all prespecified families in the Ebola sample.) The effects on general utilization remain positive and significant when we pool the treatments and in CM alone; the effect is attenuated in the NFA arm relative to the full sample. We continue to find positive effects on satisfaction, focusing here on satisfaction with public health workers, which is asked of all households.[40] Both treatment arms generate roughly equivalent increases in satisfaction with public health workers, on the order of 0.15 standard deviations. Finally, we find improvements (about 0.10 standard deviations) in households' attitudes toward Western-style medicine, particularly its effectiveness relative to traditional healers or spiritual remedies. While this indicator is not listed among the outcomes in our analysis plan, its inclusion was motivated by assessments of the Ebola crisis stressing the importance of trust in Western-style medicine (e.g., Kruk et al. 2015). We combine these three measures into a perceived quality of care index at the household level. In the top row of Table V, we find that both CM and NFA (pooled and separately) have significant effects on this index.

To visualize the relationship between perceived quality of care and reporting, we aggregate this index to the clinic level and plot it against total cases (both variables residualized) in Figure IV. The solid line corresponds to an instrumental variables (IV) estimate of the effect of perceived quality of care on reported Ebola cases, with perceived quality of care instrumented by treatment. The slope is positive, suggesting that clinics with larger improvements in perceived quality of care saw larger increases in reporting.[41]

Online Appendix Table E.27 presents the IV analysis corresponding to this figure. The top panel shows the first-stage effect,

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

41. This approach is similar to Kling, Liebman, and Katz (2007), who explore whether the Moving to Opportunity program affects individual outcomes through its effect on neighborhood poverty; we thank Lawrence Katz for suggesting it.

TABLE V

PERCEIVED QUALITY OF CARE

	Control mean (1)	Pooled (2)	CM (3)	NFA (4)	Difference (5)	Joint F-test (p) (6)	N (7)
Perceived quality of care	0.000	0.198	0.221	0.180	0.041	8.696	3,178
	(1.000)	(0.048)***	(0.062)***	(0.052)***	(0.062)	(.000)***	
General utilization	0.000	0.098	0.135	0.070	0.065	2.792	2,085
	(1.000)	(0.045)**	(0.058)**	(0.047)	(0.053)	(.064)*	
Satisfaction with public health workers	0.000	0.151	0.150	0.152	− 0.002	4.807	3,149
	(1.000)	(0.049)***	(0.065)**	(0.053)***	(0.064)	(.009)***	
Relative effectiveness of western medicine	0.000	0.102	0.117	0.090	0.026	2.040	2,663
	(1.000)	(0.051)**	(0.064)*	(0.054)*	(0.060)	(.133)	

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

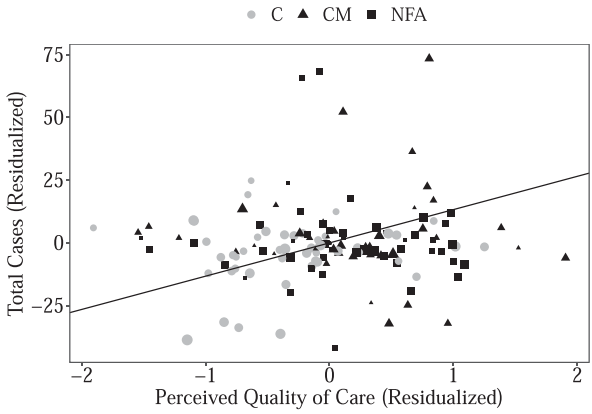


FIGURE IV

Visualizing the Relationship between Perceived Quality of Care and Ebola Cases

Figure uses cross-sectional data from 160 sections. We first residualize both the perceived quality of care index (x-axis) and total cases (y-axis) using matching-triplet and week fixed effects, as well as the baseline value perceived quality of care index. We plot the residualized values, with dots sized according to the weight they receive in the regression. The solid line corresponds to an instrumental variables estimate of the effect of perceived quality of care on reported Ebola cases, with perceived quality of care instrumented by treatment. This IV analysis is shown in Online Appendix Table E.27.

and the bottom panel the IV estimate: we find that a one standard deviation change in the perceived quality of care corresponds to an increase in weekly case reports of 0.39 cases per section. We recover nearly identical IV estimates when we use CM and NFA as separate instruments (in the second column of Online Appendix Table E.27). The first-stage effect is larger in CM, owing to its larger effect on general utilization in this subsample. (The larger first-stage effect for CM aligns with our results in Table III where we see larger, if statistically indistinguishable, effects of CM on total cases.) Note, however, that this approach quantifies the effect of quality of care on reported Ebola cases only under the strong assumption that the entire effect of the instrument works through changes in the perceived quality of care. We view these results as suggestive.

We also find that endogenous changes in the perceived quality of care are associated with greater reporting of total cases in control sections. In Online Appendix Table E.28, we regress total cases in control sections on the change in the perceived

quality of care index between baseline and endline. We find a positive relationship, which is significant at the 10% level when we control for population and include fixed effects for week and chiefdom, the administrative unit just above sections. The magnitude of the correlational relationship is about half as large as that estimated using 2SLS, suggesting that it understates the effect of quality of care on reporting of Ebola cases.

As with the full sample, we do not find consistent positive effects for families focused on supply-side variables in the Ebola sample. Pooling the treatments, we see no significant effects on health service delivery or clinic organization and services (see Online Appendix Table E.26).[42]

IV.C. Cost-Effectiveness

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

We pit these interventions against a well-regarded but reactive approach to Sierra Leone's Ebola crisis. 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 low cost among the emergency response centers (as compared to the EHCs and ETUs, which both provided intensive care and treatment). For example, CCCs were typically set up in tents or repurposed buildings and did not require new construction.

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

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 Online Appendix E.26 for details). Comparing the estimated effect size to the cost for each intervention shows that CCCs increased testing at a cost of $38,232 per case. In comparison, the pooled intervention cost only $1,084 per case. For confirmed cases, the numbers are $161,115 and $3,188, respectively.

Whether the pooled interventions or the emergency CCCs are more cost-effective in managing epidemic outcomes depends on the likelihood of an epidemic such as Ebola breaking out. Absent an epidemic, no money is spent on reactive measures, like CCCs; and the interventions incur costs without contributing to containment. Comparing the ratios of cost and effect sizes implies that the interventions are more cost-effective than CCCs for epidemic events with >2%–3% probability of occurring (see Online Appendix E.26).[43]

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.

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

V. Conclusion

We use a randomized experiment completed less than a year before the Ebola outbreak in Sierra Leone to test the effectiveness of two interventions that harness social incentives and promote accountability: one implemented community monitoring of government-run health clinics, and the other conferred status awards to clinic staff. Our findings suggest that these interventions can boost the perceived quality of health care and improve health outcomes in a developing country setting—not only during "normal" times but also during crises.

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

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

We find that both interventions substantially increased reporting of Ebola cases, by 62%. In addition, analogous to the pre-Ebola period, CM also improved health outcomes during the crisis, reducing Ebola patient mortality conditional on cases. These results suggest that improvements in the perceived quality of care at intervention clinics led to increased reporting during the crisis, and improvements in administered care in CM clinics also persisted into the crisis period. CM has qualitatively stronger effects than NFA both before the crisis, and during the Ebola outbreak; this suggests that involving the community in promoting accountability may be especially effective in improving the quality of health services. One possible reason for this effect is that public meetings act as coordination devices where community members can align beliefs and perceptions about the clinic.

We find no support for two alternative explanations of why the interventions could have increased reported case counts—by unintentionally increasing transmission or enabling more top-down surveillance. Inconsistent with increased transmission at treated clinics, we observe increased reporting by both individuals

who tested positive and those who tested negative, with no observed changes in the ratio of positive to negative case types. We also see no indication of more Ebola-specific treatment facilities, lab resources, or caseworkers in treated areas, suggesting that resources for screening and contact tracing were not targeted to areas that received the interventions.

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

University of California, Los Angeles, Luskin School of Public Affairs, United States
University of Chicago, Harris School of Public Policy, and National Bureau of Economic Research, United States
Stockholm University, Max Planck Institute for Collective Goods, and Institute for Industrial Economics, Sweden
University of California, Berkeley, Center for Effective Global Action, United States
Wageningen University and Research, Netherlands

Supplementary Material

An Online Appendix for this article can be found at *The Quarterly Journal of Economics* online (qje.oxfordjournals.org).

Data Availability

Data and code replicating the tables and figures in this article can be found at Christensen et al. (2021), in the Harvard Dataverse, doi: 10.7910/DVN/YEH04R.

References

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

Agüero, Jorge M., and Trinidad Beleche, "Health Shocks and Their Long-Lasting Impact on Health Behaviors: Evidence from the 2009 H1N1 Pandemic in Mexico," *Journal of Health Economics*, 54 (2017), 40–55.

Alsan, Marcella, and Marianne Wanamaker, "Tuskegee and the Health of Black Men," *Quarterly Journal of Economics*, 133 (2018), 407–455.

Anderson, Michael L., "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 (2008), 1481–1495.

Andrabi, Tahir, Jishnu Das, Asim I. Khwaja, Selcuk Ozyurt, and Niharika Singh, *Upping the Ante: The equilibrium Effects of Unconditional Grants to Private Schools* (Washington, DC: World Bank, 2018).

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

Ashraf, Nava, Oriana Bandiera, Edward Davenport, and Scott S. Lee, "Losing Prosociality in the Quest for Talent? Sorting, Selection, and Productivity in the Delivery of Public Services," *American Economic Review*, 110 (2020), 1355–1394.

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

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

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

Banerjee, Abhijit V., Rukmini Banerji, Esther Duflo, Rachel Glennerster, and Stuti Khemani, "Pitfalls of Participatory Programs: Evidence from a Randomized Evaluation in Education in India," *American Economic Journal: Economic Policy*, 2 (2010), 1–30.

Banerjee, Abhijit, Angus Deaton, and Esther Duflo, "Health Care Delivery in Rural Rajasthan," *Economic and Political Weekly*, 39 (2004), 944–949.

Banerjee, Abhijit V., Esther Duflo, and Rachel Glennerster, "Putting a Band-Aid on a Corpse: Incentives for Nurses in the Indian Public Health Care System," *Journal of the European Economic Association*, 6 (2008), 487–500.

Barr, Abigail, Frederick Mugisha, Pieter Serneels, and Andrew Zeitlin, "Information and Collective Action in Community-Based Monitoring of Schools: Field and Lab Experimental Evidence from Uganda," Georgetown University Working paper, 2012.

Bellemare, Marc F., and Casey J. Wichman, "Elasticities and the Inverse Hyperbolic Sine Transformation," *Oxford Bulletin of Economics and Statistics*, 82 (2020), 50–61.

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

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

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

Björkman Nyqvist, Martina, Andrea Guariso, and David Yanagizawa-Drott, "Reducing Child Mortality in the Last Mile: Experimental Evidence on Community Health Promoters in Uganda," *American Economic Journal: Applied Economics*, 11 (2019), 155–192.

Björkman, Martina, and Jakob Svensson, "Power to the People: Evidence from a Randomized Field Experiment on Community-Based Monitoring in Uganda," *Quarterly Journal of Economics*, 124 (2009), 735–769.
Blair, Robert A., Benjamin S. Morse, and Lily L. Tsai, "Public Health and Public Trust: Survey Evidence from the Ebola Virus Disease Epidemic in Liberia," *Social Science and Medicine*, 172 (2017), 89–97.
Casey, Katherine, Rachel Glennerster, and Edward Miguel, "Reshaping Institutions: Evidence on Aid Impacts Using a Preanalysis Plan," *Quarterly Journal of Economics*, 127 (2012), 1755–1812.
CDC, "Cost of the Ebola Epidemic," 2019, https://www.cdc.gov/vhf/ebola/history/2014-2016-outbreak/case-counts.html?CDC_AA_refVal=https%3A%2F%2Fwww.cdc.gov%2Fvhf%2Febola%2Foutbreaks%2F2014-west-africa%2Fcase-counts.html.
Christensen, Darin, Oeindrila Dube, Johannes Haushofer, Bilal Siddiqi, and Maarten Voors, "Community-Based Crisis Response: Evidence from Sierra Leone's Ebola Outbreak," *AEA Papers and Proceedings*, 110 (2020), 260–264.
Christensen, Darin, Oeindrila Dube, Johannes Haushofer, Bilal Siddiqi, and Maarten Voors, "Replication Data for: 'Building Resilient Health Systems: Experimental Evidence from Sierra Leone and the 2014 Ebola Outbreak'," (2021), Harvard Dataverse, doi: 10.7910/DVN/YEH04R.
Das, Jishnu, Alaka Holla, Aakash Mohpal, and Karthik Muralidharan, "Quality and Accountability in Health Care Delivery: Audit-Study Evidence from Primary Care in India," *American Economic Review*, 106 (2016), 3765–3799.
Deaton, Angus, *The Great Escape: Health, Wealth, and the Origins of Inequality* (Princeton, NJ: Princeton University Press, 2013).
Denney, Lisa, and Richard Mallett, "Mapping Sierra Leone's Plural Health System and How People Navigate it," Technical Report, ODI, 2014.
Dixit, Avinash, "Incentives and Organizations in the Public Sector: An Interpretative Review," *Journal of Human Resources*, 37 (2002), 696–727.
Dupas, Pascaline, "Health Behavior in Developing Countries," *Annual Review of Economics*, 3 (2011), 425–449.
Dupas, Pascaline, and Edward Miguel, "Impacts and Determinants of Health Levels in Low-Income Countries," in *Handbook of Economic Field Experiments,* vol. 2 eds. Esther Duflo and Abhijit Banerjee (Amsterdam: Elsevier, 2017). 3–93.
Elston, J. W. T., A. J. Moosa, F. Moses, G. Walker, N. Dotta, R. J. Waldman, and J. Wright, "Impact of the Ebola Outbreak on Health Systems and Population Health in Sierra Leone," *Journal of Public Health*, 38 (2016), 673–678.
Enserink, Martin, "How Many Ebola Cases Are There Really?," *Science Magazine,* October 20, 2014. https://www.sciencemag.org/news/2014/10/how-many-ebola-cases-are-there-really.
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, "Transmission Dynamics of Ebola Virus Disease and Intervention Effectiveness in Sierra Leone," *Proceedings of the National Academy of Sciences*, 113 (2016), 4488–4493.
Fiala, Nathan, and Patrick Premand, "Social Accountability and Service Delivery: Experimental Evidence from Uganda," World Bank Policy Research Working Paper, 2018.
Finan, Frederico, Benjamin A. Olken, and Rohini Pande, "The Personnel Economics of the Developing State," *Handbook of Economic Field Experiments*, vol. 2, eds. Abhijit Vinayak Banerjee and Esther Duflo (Amsterdam: North-Holland, 2017), 467–514.
Glewwe, Paul, Nauman Ilias, and Michael Kremer, "Teacher Incentives," *American Economic Journal: Applied Economics*, 2 (2010), 205–227.
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, "Variability in Intrahousehold Transmission of Ebola Virus, and Estimation of the Household Secondary Attack Rate," *Journal of Infectious Diseases*, 217 (2018), 232–237.

Greevy, Robert, and Cole Beck, "nbpMatching Demo: Triplet Matching Prior to Randomization," (2016), https://cran.r-project.org/web/packages/nbpMatching/nbpMatching.pdf.

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

Kling, Jeffrey R., and Jeffrey B. Liebman, "Experimental Analysis of Neighborhood Effects on Youth," Working Paper, Harvard University, 2004.

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

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

Kruk, Margaret E., Anna D. Gage, Catherine Arsenault, Keely Jordan, Hannah H. Leslie, Sanam Roder-DeWan, Olusoji Adeyi, and Pierre Barker et al., "High-Quality Health Systems in the Sustainable Development Goals Era: Time for a Revolution," *Lancet Global Health*, 6 (2018), e1196–e1252.

Kruk, Margaret E., Michael Myers, S. Tornorlah Varpilah, and Bernice T. Dahn, "What Is a Resilient Health System? Lessons from Ebola," *Lancet*, 385 (2015), 1910–1912.

Lautharte, Ildo, Jr., and Imran Rasul, "The Anatomy of a Public Health Crisis: Household and Health Sector Responses to the Zika Epidemic in Brazil," Working Paper, 2019, https://www.ildolautharte.com/publication/zika/.

Levy, Benjamin, Carol Y. Rao, Laura Miller, Ngozi Kennedy, Monica Adams, Rosemary Davis, Laura Hastings, Augustin Kabano, Sarah D. Bennett, and Momodu Sesay, "Ebola Infection Control in Sierra Leonean Health Clinics: A Large Cross-Agency Cooperative Project," *American Journal of Infection Control*, 43 (2015), 752–755.

Lowes, Sara, and Eduardo Montero, "The Legacy of Colonial Medicine in Central Africa," Working Paper, 2018.

Maffioli, Elisa M., "The Political Economy of Health Epidemics: Evidence from the Ebola Outbreak," IGC Working Paper No. S-51208-LIB-1, 2018.

Mansuri, Ghazala, and Vijayendra Rao, "Localizing Development: Does Participation Work?" World Bank Technical Report, 2003.

Markham, Steven E., K. Dow Scott, and Gail H. McKee, "Recognizing Good Attendance: A Longitudinal, Quasi-Experimental Field Study," *Personnel Psychology*, 55 (2002), 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, "Ebola Surveillance—Guinea, Liberia, and Sierra Leone," *MMWR Supplements*, 65 (2016), 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, "Innovation to Confront Ebola in Sierra Leone: The Community-Care-Centre Model," *Lancet Global Health*, 3 (2015), e361–e362.

Miller, Grant, Renfu Luo, Linxiu Zhang, Sean Sylvia, Yaojiang Shi, Patricia Foo, Qiran Zhao, Reynaldo Martorell, Alexis Medina, and Scott Rozelle, "Effectiveness of Provider Incentives for Anaemia Reduction in Rural China: A Cluster Randomised Trial," *BMJ*, 345 (2012), e4809.

MOH, "Human Resources for Health Strategy 2017–2021," Sierra Leone Ministry of Health, 2017, https://www.afro.who.int/sites/default/files/2017-05/hrhstrategy2017.pdf.

Mohanan, Manoj, Vikram S. Rajan, Kendal Swanson, and Harsha Thirumurthy, "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, 2020.

Morse, Ben, Karen A. Grépin, Robert A. Blair, and Lily Tsai, "Patterns of Demand for Non-Ebola Health Services during and after the Ebola Outbreak: Panel Survey Evidence from Monrovia, Liberia," *BMJ Global Health*, 1 (2016), e000007.

Olken, Benjamin A., "Monitoring Corruption: Evidence from a Field Experiment in Indonesia," *Journal of Political Economy*, 115 (2007), 200–249.

Olken, Benjamin A., Junko Onishi, and Susan Wong, "Should Aid Reward Performance? Evidence from a Field Experiment on Health and Education in Indonesia," *American Economic Journal: Applied Economics*, 6 (2014), 1–34.

Owada, Kei, Tim Eckmanns, Kande-Bure O'Bai Kamara, and Olushayo Oluseun Olu, "Epidemiological Data Management during an Outbreak of Ebola Virus Disease: Key Issues and Observations from Sierra Leone," *Frontiers in Public Health*, 4 (2016), 163.

Özler, Berk, Lia C. H. Fernald, Patricia Kariger, Christin McConnell, Michelle Neuman, and Eduardo Fraga, "Combining Pre-School Teacher Training with Parenting Education: A Cluster-Randomized Controlled Trial," *Journal of Development Economics*, 133 (2018), 448–467.

Pradhan, Menno, Daniel Suryadarma, Amanda Beatty, Maisy Wong, Armida Alishjabana, Arya Gaduh, and Rima Prama Artha, *Improving Educational Quality through Enhancing Community Participation: Results from a Randomized Field Experiment in Indonesia* (Washington, DC: World Bank, 2011).

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

Raffler, Pia, Daniel N. Posner, and Doug Parkerson, "The Weakness of Bottom-Up: Experimental Evidence from the Ugandan Health Sector," Working Paper, Harvard University, 2019.

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

Stephenson, N., K. Miller, M. Gallivan, C. Lam, V. Serhiyenko, and N. Madhav, "Filovirus Model Catalog V2," May 28, 2020.

Tsai, Lily, Benjamin Morse, and Robert Blair, "Building Trust and Cooperation in Weak States: Persuasion and Source Accountability in Liberia during the 2014–2015 Ebola Crisis," Working Paper, 2019.

UNICEF, "Sierra Leone Health Facility Survey 2014: Assessing the Impact of the EVD Outbreak on Health Systems in Sierra Leone," UNICEF Technical Report, 2014.

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, "Ebola and Community Health Worker Services in Kenema District, Sierra Leone: Please Mind the Gap!," *Public Health Action*, 7 (2017), S55–S61.

Vinck, Patrick, Phuong N. Pham, Kenedy K. Bindu, Juliet Bedford, and Eric J. Nilles, "Institutional Trust and Misinformation in the response to the 2018–19 Ebola Outbreak in North Kivu, DR Congo: A Population-Based Survey," *Lancet Infectious Diseases*, 19 (2019), 529–536.

Wen, Leana S., "Governments Need People's Trust to Stop an Outbreak. Where Does that Leave Us?" *Washington Post*, January 22, 2020.

WHO, "Experimental Therapies: Growing Interest in the Use of Whole Blood or Plasma from Recovered Ebola Patients (Convalescent Therapies)," WHO Technical Report, 2014, https://www.who.int/mediacentre/news/ebola/26-september-2014/en/.

———, "Clinical Management of Patients with Viral Haemorrhagic Fever," WHO Technical Report, 2016, https://www.who.int/csr/resources/publications/clinical-management-patients/en/.

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