Managerial Practices and Altruism in Health Care Delivery*

David Contreras-Loya Paul J. Gertler Ada T. Kwan

April 5, 2021

Abstract

Managerial ability influences firm performance in various industries. This suggests that improving business performance through better management in health care can result in productivity gains in countries with low and varied quality of care. We report results from a field experiment of a large and comprehensive management consulting intervention designed to improve business management and care delivery in the Kenyan private health sector. We find large improvements in management practices and structural quality that translated into better business performance in terms of increased investment, output, higher prices, increased revenue, lower unit costs and higher profits. However, better management and structural quality did not translate into improved process (clinical) quality. In fact, we surprisingly find the program significantly reduced correct clinical case management of patient care by 12% (p-value = 0.021). We also find that the fall in quality did not register with clients and did not affect demand, consistent with demand being quality inelastic. Hence, it was optimal for profit-maximizing firms to lower quality and raise prices. We examine this further by measuring provider-specific preferences with a modified dictator game: we find that the least altruistic (most profit maximizing) providers in the program were the ones that reduced correct clinical case management, while charging 121% more than the least altruistic providers outside the program. Altruistic providers in the program did not lower quality or raise prices.
JEL Codes: D40, I11, L2, O12.

*Contreras-Loya: School of Public Health, UC Berkeley (email: davidcon@berkeley.edu); Gertler: Haas School of Business, UC Berkeley and NBER (email: gertler@berkeley.edu); Kwan: Department of Medicine, UCSF (email: ada.kwan@ucsf.edu). We are grateful to Claire Boone, Rita Cuckovich, Joshua Gruber, Afke Jager, and Nicole Perales for their research assistance. We thank Innovations for Poverty Action in Kenya for their fieldwork management; our Technical Advisory Group for their expertise; individuals at Marie Stopes International, Population Services International, PharmAccess Foundation, UK DFID, and BMGF for their comments to previous versions of these results. The authors thank Catherine Goodman, Jishnu Das, Pamela Jakiela, Jessica King, Mylène Lagarde, Jing Li, Timothy Powell-Jackson, Kara Hanson, Dominic Montagu, the participants of the 10th SME Working Group Meeting, UC Berkeley Development Seminar, and the 2020 Empirical Management Conference for their discussions, feedback, and comments. This study was funded by DFID and BMGF.

1 Introduction

Private for-profit health care providers are important players in the health care sector in many low- and middle-income countries (LMICs) and have the potential to provide high-quality, affordable health care services to the 7.5 billion people worldwide who currently do not have access to essential health services (Evans et al., 2001). For example, the private for-profit health care sector in Kenya is as large as the public sector in terms of number of facilities and account for about one-third of preventive and curative patient care.[1]

These private health care markets are comprised of many small firms that are managed by clinicians with little or no training in business management or continuous quality improvement (Barnes et al., 2010). In Kenya, virtually all clinics (98%) were managed by a medical professional (doctor, clinical officer, nurse, or midwife) with little business training or background; only 26% of clinics had used a bank account that was exclusive to the clinic to manage financial transactions; and only 25% had ever taken a business loan(Gertler et al., 2020).

The lack of management skills is concerning. The literature has documented the existence of a *core* set of management practices that correlate with firm performance: standardizing operations, performance monitoring, goal setting, and incentivizing employees (Bloom et al., 2013). The idea that managerial ability influences firm performance dates back to Walker (1887), who helped to establish the notion that entrepreneurs drive industry development and jobs creation. Recent experimental evidence supports the notion that management consulting can increase productivity and profitability of both large and small firms by overcoming information constraints (Bloom et al., 2013, 2018; Bruhn et al., 2018; Karlan and Valdivia, 2011). Across nine countries, hospitals located closer to universities offering education in both medicine and business have been found to have lower mortality rates, better managerial practices and more managers with business training (Bloom et al., 2020). These productivity gains are particularly appealing to organizations operating in environments in which resources are limited and markets are highly saturated, where cost containment strategies are key for business survival.

There are considerable opportunities to improve business performance through better management in health care. Private clinics tend to be small and offer a limited range of services. In Kenya, private clinics treated an average of 67 clients per week, or about 10 clients per day (Gertler et al., 2020). This implies substantial excess capacity: that is, clinicians waiting around for clients. Increasing utilization could be expanded through outreach

[1] As of 2020, there are 12,574 operational health facilities in the country, of which 42% are private for-profit. Source: Ministry of Health, Kenya Master Health Facility List accessed on October 15, 2020 at http://kmhfl.health.go.ke/.

programs and other forms of marketing, as well as by becoming accredited to accept patients through the Government of Kenya's National Hospital Insurance Fund (NHIF), an institution from the Government of Kenya with a mandate to provide health insurance to all Kenyans. Moreover, there could be gains from introducing simple professional management practices that are still lacking among private health care providers. Activities such as financial audits, development of marketing and business plans, or systems to analyze costs or profits are virtually absent from most small, private health clinics (Barnes et al., 2010).

With respect to services received by patients, there is substantial scope to improve quality of care in the private sector. In most low- and middle-income countries, quality of care is low and variable (Currie et al., 2011; Daniels et al., 2017; Das et al., 2012, 2015; Kwan et al., 2018; Mohanan et al., 2015; Sylvia et al., 2014). In Nairobi, for example, Daniels et al. (2017) found only 52% of clinical cases were correctly managed by providers and essential actions for correct management varied widely. Nairobi has the highest living standards in the country and likely has the highest levels of quality of care in the country.

In this sense, management practices may be critical to improving productivity in health care. In health care, productivity refers to an organization's ability to translate structural inputs (such as medical equipment) and clinical knowledge into process quality (such as correct diagnosis and treatment) and health outcomes (such as mortality and morbidity). There is weak evidence of a correlation between health care inputs and processes, but there is strong evidence of a correlation between management practices and clinical and financial outcomes (Bloom et al., 2014; Das and Hammer, 2014). However, the evidence of this relationship is observational and has been limited to hospital settings in middle-high and high-income countries, while experimental evidence focuses on management of clinical care (Bloom et al., 2014; Dunsch et al., 2017).

In this study, we use a field experiment to evaluate the effect of a comprehensive management consulting intervention for private for-profit health care clinics on quality of care and business success in Kenya. Management in health care is inherently different than other industries previously studied for a number of reasons. First, health care services are highly heterogeneous and tailored to each patient. Second, a patient's service package is not chosen by the firm but rather by the specific clinician. Third, the clinic (firm) provides both the set of services from which the clinician can choose and the rules by which the clinician can make those decisions. Moreover, clinicians may have altruistic preferences that divert potential surplus towards serving more patients at higher levels of quality at the expense of profits.

We examine the effects of the African Health Markets for Equity (AHME) program, a comprehensive and ambitious intervention designed to improve business management and clinical decision making and implemented between 2013 and 2019. We use data from baseline to confirm balance across the N=232 clinics randomized to either receive the AHME program (AHME treatment) or no program (AHME control), and we examine endline data collected from a unique set of surveys[2] to estimate AHME's causal effects on quality and management outcomes. Overall, we find large improvements in management practices of both quality of care and business activities. We also find large increases in structural quality (i.e., the equipment and supplies necessary to deliver high quality care). We then examine whether the improvement in management translated into better quality of care and business outcomes.

We measure process quality (i.e., the clinical content of a visit for a particular health condition), using standardized patient (SP) audits (Kwan et al., 2019). SPs are healthy individuals recruited and trained to portray pre-designed health scenarios at sampled health facilities. The SP method is the state-of-the-art method, increasingly considered the gold standard, for assessing provider practice during a one-time interaction with a health care provider. Because providers are visited by the "same patient" with the "same illness", the typical confounders arising from differential patient and case-mix are better controlled for than in other data used to assess quality.

In terms of process quality, we surprisingly find that the AHME program significantly *reduced* correct care for outpatient services by 12% (p-value = 0.021) relative to the AHME control group mean of 63.6%. We eliminate the hypotheses that these effects can be attributed to inadequate clinical knowledge or stock levels for necessary supplies: average knowledge of how to correctly manage cases among the same providers were very high (90% for diarrhea; 98% for malaria), and in contrast to other studies, we find very low stockout rates (1%-2%) among clinics for medicines or lab tests associated with correct care.

Clients (patients) did not seem to recognize the reduction in process quality. We find very few program effects on client satisfaction and client perceptions from both SP and patient exit interview data. We further do not find any evidence that households perceived any differences in quality between AHME treatment and control clinics, meaning households in clinic catchment areas did not notice the reductions of correct care (or increases in structures) provided by AHME clinics.

One of the key ideas underlying AHME's quality improvement strategy is that, by providing clinics with the abilities to control the quality services they deliver and to manage their services efficiently, private clinics can take advantage of market opportunities to deliver quality care at financially viable prices. Central to this idea is that clients are able to observe and judge quality and are willing to pay for quality: that is, more clients will seek care at

[2] AHME endline surveys administered between 2018-2019) include: Household surveys, clinic mapping, client exit interviews, standardized patients, clinic survey, and in-charge and provider surveys, which included provider vignettes and a modified dictator game.

clinics that provide higher quality care and are willing to pay more for that care. Thus, the success of this strategy depends on the sensitivity (elasticity) of client demand to quality. However, the health economics literature typically considers demand quality inelastic based on the realization that clients have trouble assessing quality, resulting in a classic market failure of asymmetric information - an idea that has been around since Arrow (1963). If clients are not able to judge quality and providers are not rewarded for supplying quality, then a profit-maximizing clinic would use their improved control to reduce levels of quality and, if possible, to also raise prices.

Many of these predictions have been validated by the results of our impact evaluation. First, we find substantial improvement in structural quality and quality management processes due to the AHME package of interventions. Second, we observe a reduction in process quality, specifically in correct case management, which we define as the minimal and essential actions benchmarked against national guidelines. Third, client perceptions of quality and client satisfaction did not fall. Fourth, providers raised prices charged to clients.

The reductions in care quality and higher prices observed on average may mask heterogeneity driven by differences in social preferences among providers. That is, not all providers tend toward altruistic preferences, and not all providers tend toward the opposite extreme: profit maximizing preferences. The tendency to reduce quality under conditions of pure profit maximization and asymmetric information might be offset to the extent by which altruistic health professionals are willing to sacrifice profit to deliver higher quality care to their clients. This begs the question: *Do the effects of AHME on correct case management and prices charged vary according to the variation in a provider's profit-driven versus altruistic preference?*

We tested whether provider preferences for profits versus altruism play a role in the effect of AHME on process quality and prices. We elicited individual provider altruistic preferences (for providers who saw SPs) with a modified dictator game, a standard and well-validated measurement methodology in economics. We find that the least altruistic providers — that is, "the profit maximizers" — at AHME treatment clinics were the ones to reduce correct care, meanwhile charging 121% higher prices than the least altruistic providers at AHME control clinics. In contrast, the quality of care delivered and prices charged by the most altruistic providers in AHME treatment clinics were no different from those in control clinics. In other words, we find that all of the negative effect of AHME on process quality and the positive effect of AHME on prices is concentrated among profit maximizing providers, while AHME had no effect on quality or prices among the most altruistic providers.

Despite our findings on quality of care, we find clinics improved their business outcomes in several manners. First, we find substantial improvement in business management practices.

Second, on average, treatment clinics had 34% higher revenues (p-value = 0.042) relative to control clinics. We find evidence that this increase in revenues is due to both an increase in volume of clients and an increase in prices charged to clients: treatment clinics increased their patient volume by 26% (p-value = 0.056) relative to the AHME control group mean of 367 clients per month and by about 15% higher prices overall for preventive and curative health care services. We attribute the increase in patients to the combined effect of outreach, participation of community health workers (CHWs) for demand generation, branding, and empanelment with NHIF. Moreover, we find a reduction in average costs driven by both improvements in efficiency and a reduction in quality. All of this resulted in substantial improvements in profitability for clinics receiving the program.

This research makes several contributions to the literature. First, this study contributes to the growing literature on management of firms, particularly for management consulting interventions. A recent review found that business training has significant impacts on the adoption of management practices, but not on firm performance. However, these studies suffer from lack of statistical power and examine training programs of short duration (McKenzie, 2020; McKenzie and Woodruff, 2014). In the context of the Kenyan private health sector, we examine an at-scale management consulting intervention that took place for 6 years and find that despite improvements in management practices, structural quality, and business outcomes, average quality dropped due to the program. Further, among the most altruistic health care providers, the program had no effects on correct care and prices, but profit maximizing health care providers reduced quality of care and raised prices.

Second, this study contributes to our understanding of the complexities involved in improving quality at a time when stakeholders are committed to universal access to high quality care. We examine the impact of a comprehensive package of interventions aimed to improve access to quality of care across Kenya's private sector, and we are able to - for the first time across the same set of clinics to our knowledge - explore quality improvement and the roles of client satisfaction and perceptions of client amenities, as well as provider competence, practice, and preferences.

Third, this study methodologically contributes to the literature on quality of health care measurement. Most research studies on health services exclude multiple quality of care measures and/or suffer from endogeneity problems when estimating quality. This is often because of measurement and data issues, as well as budget constraints. Together, these findings have implications for the growing literature on information asymmetries and uncertainty in the context of insurance, health markets, and the production of health services (Fitzpatrick and Tumlinson, 2017; Gertler and Waldman, 1992; Hurley, 2000).

The remainder of this paper is as follows. Section 2 describes the AHME program, and section 3 describes our methods, including the AHME impact evaluation's experimental design, AHME program compliance, our measures of interest and surveys used to collect a unique set of data, and our empirical approach. Sections 4 and 5 provides our results and discussion, respectively.

2 The AHME Program

The AHME program identified 3 necessary conditions for health markets to work well: (1) primary health care must be covered by national health insurance and clinics must accept national health insurance, (2) high quality services must be available through those clinics, and (3) clinics must have viable financial business models. AHME core activities built on social franchising carried out by Marie Stopes Kenya (MSK) under the Amua brand and Population Service Kenya (PSK) under the Tunza brand. Once clinics participating in AHME joined one of the franchising networks, they were also enrolled in SafeCare, which provided supplemental guidance on improving and maintaining quality of care, as well as business support through a program that helped clinics upgrade management practices. In addition, clinics were offered optional help with accreditation for NHIF. Each of these interventions in the AHME package are described below.

1. **Social Franchising:** Social franchising binds together a large number of small, independent providers (clinics) in order to leverage scale to improve supply chains, exploit joint brand advertising, and strengthen worker training and supervision. Social franchises are able to scale rapidly, decrease transaction costs, standardize services, collectively negotiate financial reimbursement mechanisms, and replicate best practices among a large group. Brand advertising and education programs help promote franchise products and strengthen critical links to low-income consumers. In Kenya, franchising for the AHME program was overseen by two social franchise networks: MSK's Amua network and PSK's Tunza network. In order to be franchised, clinics were required to meet minimum standards, including proper licensing, no memberships in other networks, regular and adequate operating hours, as well as other requirements for services offered. The specific services[3] offered as part of franchising focused on: expanded scope of clinical services, improved quality through training for clinical staff, increased clinical quality, increased demand for clinic services, subsidized commodities and equipment, and monitoring and reporting.

[3] See Appendix A for further details of franchising services.

2. **SafeCare** is a certification program that focuses on improving the environment in the clinic in order to facilitate the delivery of high quality services. It complements the in-depth quality standards and training addressed through franchising. Accredited by the International Society for Quality in Health Care (ISQua), SafeCare's primary goal is to "increase trust between patients, providers, financers, and governments by making medical and financial risks more transparent and identifying quality gaps, paving the way for sustainable investment in quality, required to improve scale and scope of health care delivery in lower and middle income countries" (PharmAccess Group, 2017). To achieve this, the SafeCafe program focused on the following 13 service areas: governance and management, human resource management, patient and family rights and access to care, management of information services, risk management, primary health care services, inpatient care, surgery and anesthesia services, laboratory services diagnostic imaging services, medication management, clinic management services, and support services. During the AHME program, clinics were first assessed across various service areas (PharmAccess, 2017a,b) from which MSK and PSK developed customized Quality Improvement Plans (QIP) for each clinic (PharmAccess, 2017c,d). Subsequently, implementing partners conducted routine visits at clinics to check whether significant progress had been made on the QIP, and this was followed by a reassessment and QIP revision. Further details on the service areas are detailed in Appendix A.

3. **Business Support:** The main aim of Business Support in AHME was to enable owners to run their clinics as profitable professional businesses. The franchisers first assessed clinic business practices using a Business Assessment tool (Population Services Kenya, 2016), from which they helped each clinic develop a Business Improvement Plan (BIP) (AHME Program, 2016). Franchisers organized multi-day business training workshops and helped to set up business structures and systems, such as keeping financial records, doing stock management and developing strategic plans. The franchisers also gave guidance on business development and facilitated financing for expansion by helping obtaining bank loans. The Business Assessment tool measures the quality of clinic business structures and systems through five categories (general business operations, financial management, banking and banking records, stock management, and marketing and demand creation)[4], each containing 3 to 7 criteria. Each criterion is ranked Severe, More Severe, or Most Severe as a way to determine which item should be prioritized as necessary to improve clinic operations. A Business QIP was drawn up

[4]See Appendix A for further details

and progress for areas receiving a low score on the Business Assessment was monitored. Clinics could also participate in Business Training, a multi-day class held regionally where staff could learn best practices for business and financial management.

4. **NHIF Empanelment Support:** As part of the business support component, AHME helped clinics become accredited by NHIF in order to be able to treat insured clients and accept NHIF insurance payments for those services. The hypothesis was that increasing the number of clinics accepting NHIF insurance would increase the potential client base of clinics and improve the affordability of care. The NHIF Empanelment Support program was designed to educate clinic owners about NHIF empanelment, and assist clinics interested in empanelment throughout the application process. Prior to AHME, while clinic interest in empanelment was high and the number of clients using NHIF insurance in Kenya was increasing, the numbers of private clinics in Kenya accepting NHIF insurance was low. Only 12% of the 123 AHME treatment group clinics were accepting NHIF insurance prior to AHME. This was likely to be partially due to the bureaucratic, time-consuming process to become empaneled.

3 Methods

In this study, we analyze endline patient, provider, clinic, and household-level data we collected between 2018-2019 across n=232 primary care clinics that comprise the AHME evaluation clinic sample. The AHME evaluation clinic sample (n=232) is distributed across 35 of Kenya's 47 counties (see Figure 1).

In this section, we describe the experimental design of the AHME impact evaluation, compliance to the AHME program, notes on measurement mapped to our data sources and analytic samples for this paper (including details on balance achieved at baseline), and ethical clearance and considerations related to this study.[5]

3.1 Experimental Design

The quasi-randomized AHME impact evaluation was conducted across Kenya between 2013-2019 and consists of two main waves of data collection (conducted at baseline and at endline) and tracking data across that time period. This experimental design of the AHME impact evaluation allows us to conduct intent-to-treat (ITT) analyses to examine quality of care

[5]Additional details on the larger AHME impact evaluation are included in Appendix B, and the reader can access Gertler et al. (2020) for additional details.

effects of the program. We describe the mapping and randomization processes and program compliance below.

Before the AHME program began in 2012, we mapped all clinics in the evaluation study areas (35 of Kenya's 47 counties) with the goal of randomizing clinics eligible for the program into treatment and control groups. Clinics were first identified for mapping using various identified sources.[6] Government clinics (public clinics and hospitals), faith-based clinics (identified by clinic name), and clinics that were identified as franchised (by franchise branding on clinic exterior) were excluded. To exclude clinics that the implementing partners indicated were not eligible for franchising or the AHME set of interventions, we administered a pre-screening and baseline survey among remaining clinics. Clinics that met the basic eligibility criteria still varied in their "level of eligibility" based on their existing capacity and suitability for franchising services and AHME interventions. Using additional criteria, created in a collaborative manner with MSK and PSK, clinics were further categorized into groups based on how likely they were to be eligible ("eligibility tiers") using data collected through the baseline survey instrument. After clinics were categorized by eligibility tier, the research team conducted a stratified randomization of eligible clinics. For the randomization process, clinics were grouped into their eligibility tiers within a county based on partner-provided criteria and data from the baseline survey, randomly ordered within strata (groups within which randomization would occur), and then randomly assigned to treatment (eligible to be offered AHME franchising immediately) or control (not eligible to be offered AHME until the completion of the study).

After randomization procedures were completed, we provided MSK and PSK with partner-specific recruitment lists indicating the order in which clinics on their lists were to be approached for screening ("sensitization") and recruitment. Once randomization procedures had been completed, MSK and PSK began engaging clinics on their lists and proceeded with

[6]These included: (1) official government list of private clinics in the country, (2) clinics belonging to a major professional health association (e.g., the Kenya Nurse and Midwives Association), (3) clinics that the AHME implementation partners suggested should be visited, and (4) additional clinics identified by evaluation teams in the field during the mapping process, but not included on any of the above lists. During the mapping process, more than 7000 potential clinics were identified nationally from government lists, partner recommendations, and field team observations. From this universe of clinics, 4216 were excluded for various reasons, including already franchised, faith-based or public clinic, and proximity to an existing franchised clinic. The remaining 3467 clinics were screened by the evaluation team using standardized surveys (pre-screening and baseline clinic surveys). Of these, 472 clinics were excluded for not meeting initial screening criteria. The remaining 3002 clinics were randomized into treatment and control clinics, assigned to MSK or PSK, and randomly ordered using the stratified design. Clinics considered ineligible were removed from the sample using the original criteria prior to the first recruitment round (N = 836; treatment: N = 418; control: N = 418) and the updated criteria prior to the second round of recruitment (N = 486; treatment: N = 258; control: N = 228). The remaining 1680 randomly ordered and randomly assigned clinics were considered eligible (treatment arm: N = 851; control arm: N = 829). Recruitment lists were prepared for MSK and PSK to visit the 851 clinics in the treatment arm.

their respective recruitment procedures in October 2013. The second round of screening and recruitment by the partners served to identify clinics that were eligible for franchising and AHME interventions in the treatment arm. Eligible clinics that were invited to join either franchise ("ever franchised") were considered part of the evaluation sample in the treatment arm. Consistent with assumptions for ITT, these ever-franchised clinics were considered part of the final evaluation sample regardless of whether they completed the franchise enrollment process or maintained their franchise enrollment status for the entirety of the study period (for any reason).

The final AHME evaluation samples were identified over the course of recruitment and honing activities. In total, 232 clinics were identified for the final evaluation sample (treatment clinics: N = 123; control clinics: N = 109; see Figure 1). In September 2016, baseline data collection, including baseline household and client exit interviews, was completed for all AHME clinics.[7]

3.2 Program Compliance

To understand the extent of program compliance, we tracked which clinics received the AHME's four clinic-level interventions-social franchising, SafeCare, business support, and NHIF empanelment support-in clinics assigned to the "treatment group" (i.e., participating in the AHME program) and clinics assigned to the "control group" (i.e., not participating in the AHME program). The results of this monitoring demonstrate that implementation of the AHME package of interventions was done well at the clinic level. Approximately a quarter of clinics disenrolled from AHME over time. Among the treatment clinics who remained until the end of the program, 97% of those that were ever franchised were visited at least once a quarter by a Social Franchising Coordinator, and 74% were visited monthly or more often. At least 90% of clinics in the treatment group had exposure to Social Franchising, for more than 18 months, while 80% had SafeCare and Business Support for the same period. Approximately 76% were offered help with NHIF empanelment. An additional 15% of treatment clinics were already empaneled at the beginning of the study, and at the end of the AHME program, 41% of treatment group clinics were empaneled in NHIF. An additional 43% were taking part in the empanelment process but were not yet empaneled.

[7]Appendix B describes treatment arm recruitment and control honing.

3.3 Measures and Data Sources

3.3.1 Quality of Care

To understand the impact of AHME on quality of care, we draw from Donabedian's conceptual framework for quality of care (see Table 1) and construct metrics based on national health guidelines (Donabedian, 1966, 1978, 1988). This section describes our quality of care framework, measures, and data.

Conceptual Framework

The framework categorizes dimensions of quality of care into three measurements: *health care structures*, *processes*, and *outcomes*. Health care structures, processes, and outcomes play a role in the interaction between an individual's health and multiple actors: the client; the provider (health care practitioners, including doctors, nurses, clinical officers, and midwives); and the clinic (including the environment and resources).

Structural quality refers to the material resources, human resources, and organizational structure that facilitate the process of seeking and providing care. Typically, these inputs are favorable for good process quality, but they do not secure it. For example, having a weigh scale in the examination room improves structural quality, but it does not directly ensure that it is used properly, if used at all, to provide effective care to clients.

Process quality refers to clients seeking and receiving care, as well as providers' diagnostic and treatment actions. Typically performed by health care providers or clinic staff, process elements in medical care are conducive to better health care outcomes but do not guarantee them. For example, a medical doctor may correctly perform differential diagnosis (the process of differentiating between two or more health conditions that share similar history or symptoms) based on the client's symptoms and responses to history questions, and follow up by prescribing efficacious medicine for the client's ailment. However, if the provider also prescribes a harmful medicine, unrelated to the ailment, and the client consumes the medicine, this may sequester symptoms and delay or complicate onward improvement of the client's health condition. Process quality is influenced not only by providers' actions, but also by client and provider characteristics.

Lastly, dimensions of care quality that are related to *health care outcomes* include health status, as well as broader elements related to "psychological function or social performance," such as client characteristics, knowledge, behaviors, and satisfaction (Donabedian, 1978).

For each of the three types of quality of care measurements (process quality, structural quality, and health care outcomes), we further separate dimensions into two components: (1) management of interpersonal relationships between provider and client; and (2) technical performance of providers, including the recommendation and implementation of appropriate care strategies. We refer to these as the *interpersonal component* of quality and the *technical component* of quality, respectively.

Quality Measurement

To examine whether the AHME program had effects on structural, process, and health care outcomes, we first discuss understanding and measuring quality. Understanding quality of care is complex and challenging, and the literature on methods to measure quality of care is vast. To summarize, popular methodologies to understand quality of care include: patient exit interviews, provider vignettes, direct observation, medical record extraction, and standardized patients (SPs).

Our study methodologically relies on the state-of-the art SP method, which can provide an unbiased and scientifically valid technique to answer questions related to provider practice (a dimension of process quality). The SP method is increasingly considered the gold standard for measuring provider practice in a one-time encounter in countries where medical records are poor or absent (King et al., 2019; Kwan et al., 2019; Peabody et al., 2000). To assess provider practice through other alternative methods would subject the data and their interpretations to various biases: (i) exit interview data, derived from surveying patients upon exiting a clinic after receiving services, is subject to recall bias and endogeneity issues including patient sorting and case-mix differences across providers; (ii) direct observation data, derived from hiring an enumerator to observe provider actions during a clinical encounter with a patient, is primarily biased by the Hawthorne effect, as well as selective and/or distorted perception of the observer, who can also be placed in an unethical position of whether to interject when a provider does something clinically questionable or incorrect; (iii) provider interview or vignette data reflects provider knowledge rather than actual practice; and (iv) medical record or administrative data, which most often only offers a glimpse of how treatment is given once diagnosed, is largely missing or incomplete in LMICs (Kwan et al., 2019). The SP method is not without limitations as SP cases are limited to a particular set of health conditions or presentation of certain health services (e.g., SPs are not best suited to assess provider practice for chronic conditions or health conditions that rely on continued interactions between providers and patients). Additionally, the SP method has not yet been validated to reflect multiple and sequential visits to a provider without risking SPs being detected as not real patients. Thus, SP data reflect a one-time encounter with a provider.

To capture aspects that define health care structures, processes, and outcomes from different perspectives, we implemented the following five surveys at endline: (1) a clinic survey; (2) a standardized patients survey; (3) a provider survey; (4) client exit interviews; and (5) household surveys. The quality of care measures we collected across these surveys are described below.[8]

Structural Quality

To capture structural quality, we implemented a two-step process. First, we collected 3,178 indicators from monitoring data collected and reported by SafeCare, MSK, and PSK. Since we did not want to independently collect data on all the indicators, we applied Item Response Theory (IRT) and reduced this number to 32 internal monitoring indicators. IRT is a method for developing survey instruments and other tests that relates performance on test questions to the respondent's level of the latent trait being tested. Twenty eight of the 32 indicators measure structural quality dimensions, which we used to construct indices of structural quality and its subcomponents: interpersonal care and technical care. We captured them through clinic administrator responses and enumerator observations in the AHME endline clinic survey.

For investments in equipment, a dedicated module asked for the existence, quantity, and value of a precoded list of 33 equipment items, plus three slots for additional equipment not listed in the options. This resulted in 54 different equipment categories, which were further aggregated into the following 5 categories: amenities; basic diagnostic equipment; laboratory equipment; medical equipment; and information technology (IT) equipment. Equipment value was calculated using the question: "How much would the value (in KSH) of [EQUIPMENT] be if the clinic had to purchase/obtain it today in its current state?" We dropped all investments that occurred before 2018. We chose this time frame because treatment clinics were enrolled at different times and therefore had different exposure times, and by 2018 all clinics in the treatment group were already franchised.

Additionally, as part of structural quality, we measure provider knowledge as a structural component that is necessary but does not directly ensure process quality. For provider knowledge, enumerators administered provider surveys to clinic owners or in-charges and providers seeing outpatient clients. The provider survey collected provider details and included clinical vignettes, which are designed to capture provider knowledge of correct case management and are administered by an enumerator in an interactive interview. For example, in one case scenario, if the respondent states he would take the client's temperature, the enumerator responds with the pre-scripted temperature measurement, but unlike a client,

[8] More details on the surveys can be found in Gertler et al. (2020), including the mapping of AHME theory of change to quality of care outcomes and indicators and details on the purpose of the AHME impact evaluation's surveys, recruitment and indicators.

the enumerator is known by the provider to be a data collector, guides the provider through the interview, and also notes down responses. There were three vignettes: (1) a mother with a child at home sick with childhood diarrhea, (2) a female seeking family planning services, and (3) an individual who has malaria. Table 2 provides details of the case scenarios, and Table 3 contains how the provider knowledge of correct case management binary outcome is constructed (described more in detail in the next section).

Process Quality

To measure process quality, we collected and analyzed primary data from N=1195 standardized patient (SP) visits. Based on the SP method, four case scenarios were developed and implemented to assess quality of outpatient care for childhood diarrhea, family planning, adult asthma, and adult malaria services, and each clinic was assigned to receive at least one visit from each of these scenarios.[9] Described in table 2, all scenarios except the asthma case were implemented as a vignette so that we would be able to compare knowledge versus practice.

We designed the SP surveys to capture the remaining four (of 32) monitoring indicators (all four were identified by IRT and were process quality measures for the management of childhood illnesses). All SP visits were unannounced, and all SPs were blinded to AHME treatment status at the clinic level, our outcomes, and the definitions of our outcomes. Further details on the SP design and method are described in the appendix.

With the vignette and SP data, we constructed binary measures for correct case management, whether any valid or unnecessary laboratory tests were ordered, and whether any efficacious or non-efficacious (including harmless and harmful) medicines were prescribed or dispensed. These technical process quality measures are shown by case in table 3. Correct case management definitions refer to minimal and essential actions as benchmarked against national guidelines for case management (Kenya Ministry of Health, 2010a,b, 2014; Kenya Ministry of Public Health and Sanitation, 2010; Kenya President's Malaria Initiative, 2018). We confirmed these definitions for minimal, essential actions, as well as our classifications for valid or unnecessary lab tests and medicines, with guidance from our technical advisory group.[10]

[9]We also conducted experiments in the SP surveys that varied one characteristic in SP case presentation for each case scenario and randomly assigned them to the AHME evaluation clinic sample. Results for these experiments are not discussed in this paper, but we control for these experiments in our analyses.

[10]A technical advisory group consisting of four Kenyan clinicians advised our team on case development, and all hired SPs participated in developing standardized narratives (e.g., name, age, family situation, living situation, etc.) for each of the SP cases during training. The technical advisory group participated in SP training and advised on outcome measures for each case.

- ***Management of Childhood Diarrhea:*** For the childhood diarrhea case, the visit was coded as managed correctly if the provider gave or advised on oral rehydration salts (ORS) or referred or asked the client to return. Ordering a stool test for this case was valid, with all other tests considered unnecessary. We classified ORS and zinc to be efficacious for the childhood diarrhea case scenario and all other medicines to be non-efficacious. For the purpose of our SP experiments, the deworming drug albendazole was non-efficacious and harmless, and the antibiotic amoxicillin was non-efficacious and harmful for this case.

- ***Family Planning Counseling:*** For family planning, the case was coded as managed correctly if the provider performed all four of the following actions: asked any family planning history questions; asked any obstetric history questions; ruled out pregnancy; and asked the client her preferred family planning method. We also examined a lenient version of this definition by reporting whether one of the four actions was performed. Table 6.3 describes the components of correct case management for the family planning case scenario. A pregnancy test was valid (in order to rule out pregnancy), and all other tests were unnecessary. Contraceptive pills were considered efficacious, with all other medicines considered non-efficacious for this scenario.

- ***Management of Asthma:*** Providers were coded as correctly managing the mild asthma case if they treated the case with an inhaler or bronchodilator, such as salbutamol, cetirizine, or prednisolone—all considered to be efficacious medicines. Any other medicines were considered non-efficacious, including franol, considered efficacious for severe asthma, but not mild asthma. Any lab tests ordered for this case were coded as unnecessary.

- ***Management of Malaria:*** For malaria, providers were coded as correctly managing the case if they ordered a malaria rapid diagnostic test (RDT or mRDT) or malaria microscopy test, and not if otherwise.[11] Blood count and brucellosis tests were considered valid. A first-line malaria treatment, artemether lumefantrine (AL), and paracetamol to manage fever were considered efficacious medicines for the malaria case, based on national guidelines for malaria management.

Health Care Outcomes

[11]Kenya has malaria endemic and non-endemic regions. For this reason, it can be less of a concern that providers do not issue a test in certain situations. Based on our SP case scenario, which is scripted with travel to a malaria endemic region during Kenya's wet season (fieldwork was March-May 2019 with all malaria visits conducted in May), based on national guidelines, and based on some non-endemic regions observing high travel from around the country, we present our data as stated.

To capture data related to client experience, perceptions of amenities, and satisfaction, we asked questions to the SPs and to actual clients in exit interviews. We administered household surveys in clinic catchment areas to understand household perceptions of clinic quality.

3.3.2 Provider Altruistic Preferences

Following literature that suggest provider preferences play a role in decision making among health care providers, we examine provider preferences, specifically distributive social preferences. Distributive social preferences-including altruism and spitefulness, fairness and aversion to inequity, and concerns regarding efficiency-relate to an individual's preferences toward the distribution and magnitude of payoffs among others on a range of issues (e.g., social security, health care, benefits). In the context of the provision of quality health care in the private sector, we focus on altruistic social preferences - a type of distributive preference that we hypothesize to govern the trade-offs that a health care provider makes between his payoffs and the payoffs to his clients (Fisman et al., 2017).

To elicit altruism, we use a real-stakes modified dictator game, and in our case, the dictator is a provider at one of the clinics in the AHME evaluation sample. Dictator games provide a nonstrategic environment to elicit social preferences. We adapt a modified dictator game originally modified from Andreoni and Miller (2002) in several subsequent studies (Balakrishnan et al., 2020; Fisman et al., 2017; Jakiela, 2013; Li et al., 2017). In our game, the provider, who is given 5 different scenarios of real or fictitious clients with different characteristics (e.g., poor), determines how much of an endowment he wants to give to the client and how much he wants to keep for himself.

After confirming our data is consistent, we examine a preference parameter for altruism estimated following Andreoni and Miller (2002); Li et al. (2017): $\alpha \in [0,1]$, which denotes the range from altruistic to fair-minded to self-interested. We focus on social preference parameter α and its ability to discern providers along the spectrum of profit-oriented to altruistic preferences. We call individuals with $\alpha = 1$ ($\alpha = 0$) perfectly selfish (perfectly altruistic), as they put all weight on the payoff to self (other). Individuals with $\alpha = 0.5$ are fair-minded, since they put equal weight on payoffs to self and other. Thus $\alpha < 0.5$ reflects individuals who tend towards altruism, and $\alpha > 0.5$ reflects individuals who tend towards self-interest.

3.3.3 Business Outcomes

To capture business outcomes for clinics, revenues, costs, and output data were collected for each clinic's last full operational month at the time of the survey.Data collection occurred from September 2018 to May 2019. Findings and observations reported here are based on a sample of 187 clinics located in 35 counties across Kenya.[12]

To collect detailed information on how clinics are managed and operated, we developed a structured, standardized survey. The survey was done in two stages given its length, complexity, and cognitive load on respondents. In both surveys, the main respondent was the manager of the clinic. However, the surveys were designed such that individual modules could be administered in a different order and to different respondents. We designed the survey in this way so that, for example, managers could direct us to the lab technician in order to obtain stock levels and prices of essential laboratory inputs. We believe this design enabled us to extract the highest quality information possible from clinic staff.

Clinic survey questionnaires included data extraction forms for capturing quantitative information, particularly for the service utilization indicators. Other data sources included Ministry of Health reporting forms (such as MoH 707), electronic databases, electronic records, electronic or written medical records, logbooks, registers, receipts, stock cards, clinic accounting records, payroll records, and performance reports. All questionnaires were administered in a quiet, private place after obtaining informed consent from the clinic manager or in-charge.

Utilization

To measure the utilization of health services, data were collected retrospectively by month for the last six full operational months. From a number of sources, we collected how many clients received health care services for each of the categories described in Table 4. We then created three mutually exclusive categories for different types of health care visits: general outpatient clients (child and adult curative services), maternal and child health care services (MCH) clients (including family planning), and inpatients. We also created a separate category for the total number of clients whose services were covered partially of fully by NHIF.

Financials

Revenues, expenses, and profits were collected retrospectively. For each clinic, we collected information from its last full operational month, and the respondent was the owner,

[12] See Appendix B for details on the sample.

in-charge, or person most knowledgeable about the clinic's finances. The AHME clinic questionnaires were designed to reduce cognitive load on respondents and to minimize measurement error by using an iterative triangulation method with automatic consistency checks, based on previous work in Uganda (Anderson et al., 2019). This method aggregated different revenue sources, as well as detailed cost categories (see Table 5). Revenues were triangulated using three figures: (1) a one month recall (based on the question: "how much were the total revenues in the last full operational month?"); (2) a monthly estimate based on weekly recall (the average of the busiest, slowest, and last week's revenue, times 4.25); and (3) a monthly estimate based on daily recall (the average of the busiest, slowest, and last day's revenue, multiplied by the number of days the clinic was open in the last full operational month). Revenues received at frequencies other than monthly, such as NHIF capitation payments (an arrangement in which the Government of Kenya pays a set amount for each person enrolled in NHIF per period of time, whether or not the enrollee seeks care) were prorated to obtain a monthly equivalent. Near the end of the financials survey module, the respondent is shown the three figures, then asked to give his or her "best estimate" of revenues for the last full operational month. This process is done before collecting any data on expenses and profits.

To collected information on expenses, we developed a comprehensive list of cost categories, tailored to the operation of a typical health clinic. The full list of 16 cost categories includes but it not limited to: payroll, electricity, waste disposal, rent, and health inputs (such as drugs and lab tests). Expenses with payment frequencies aside from monthly (such as fees paid every quarter) were converted into monthly equivalents. After completion of the expenses module, we showed the respondents calculated revenues and expenses together. At this point and before mentioning profits, respondents were given the opportunity to change the value of their previous answers if the figures were off (such as implausible expenses). This process helped reduce measurement error from both the respondent side (such as recall bias, strategic response) and the enumerator side (such as entering an extra zero).

To collect information on profits, we first asked respondents for their monthly recall, and then showed them the calculated profits (calculated revenues minus calculated costs). Respondents were then able to confirm the calculated profit, or to instead give their "best estimate." Inputs the clinic received by donation were valued at their unit costs determined by local price quotes.

Staffing Levels

The clinic survey included a staff roster module, which was administered to the person with the most knowledge about human resources. For every staff member currently working at the clinic, the questionnaire collected data on tenure, position (cadre) at the clinic, type of contractual relationship with the clinic (such as permanent, eventual, volunteer), hours worked per day, days worked per week, and monthly salary, as well as any extra income in the form of per diem or allowance received in the last month. We used the salaries information to construct total payroll expenses, and the position (cadre) to analyze staffing levels.

Prices of Health Services

Data on the price of health services were collected through several surveys. The clinic survey collected price data with six mini vignettes (separate from the vignettes used to assess knowledge of correct case management): immunization, malaria, antenatal care, sexually transmitted infection, gastroenteritis, and asthma. After presenting each mini vignette, the survey asked a series of questions about the content of a service (such as which tests or medicines were recommended), the price of that service (such as how much would be charged to the client), and the cost of that service (such as how much clinics paid to purchase those inputs). We also used information captured by the standardized patients (SP) to evaluate changes in the price of services as a result of AHME.

To understand the perception of clinic price (cost of visit) among real clients we use data from the household survey. At baseline, clients were recruited from AHME treatment and control clinics, and interviewed by the field team to understand demographics, health services utilization, and well-being of client households. At endline, we asked each household head their perceptions of health clinic price and quality.

3.3.4 Construction of Economic Outcomes

Self-reported economic outcomes in small firms in low- and middle-income countries are prone to substantial measurement error that increase the chance to find no effect of AHME, when the effect is in fact true (Fafchamps et al., 2012). To account for this problem, we use a process called winsorization to reduce the influence of outliers in the estimation of treatment effects. Winsorization replaces extreme values with the value of the highest (or lowest) data point that is not considered an outlier. This technique preserves sample size and is important for statistical inference. Here we winsorized revenues, expenses, profits, and investments at the following percentiles: (1) 1 and 99; (2) 2.5 and 97.5; and (3) 5 and 95. We report treatment effects on winsorized samples in the appendix, given the robustness of the main results using the unwinsorized data.

1. *Staff costs.* We implemented a matching algorithm that replaced missing values of monthly salary using the median value of each staff cadre (such as medical officer, registered nurse, lab technician, records officer, or administrator). For employees who

worked on a contract basis we followed a similar process, conditioning the median on the subset of individuals working under contracts. This algorithm was implemented in less than 5% of the individual-level observations from the staff roster module of the clinic survey.

2. *Input costs.* In four instances, the value of laboratory inputs and medicines was missing, due to poor record keeping. To impute these values, we ran quantile regressions, where the dependent variable was total input cost and the regressors were the values of number of clients, total hours of operation, and staff quantity. We then imputed the predicted median of input cost in the four clinics with missing values.

3. *Utilization.* For clinics with two or more months of data on number of clients, the monthly average for outpatient and inpatient volume was computed with the available data. For observations with only one value of utilization (less than 10% of the sample), we used the value of the last full operational month.

3.4 Analytic Samples and Balance at Baseline

Of the 232 clinics in the final evaluation sample, 12 had closed before the endline survey (February 2019), and 14 were excluded for various reasons: unwillingness to participate in research (N = 11); clinic could not be tracked down (N = 1); clinic is now a public hospital (N = 1); clinic located in high-risk area (N = 1). Of the remaining 206 clinics visited for inclusion in the endline survey, and after explaining the research objectives and survey content to clinic managers, 199 consented to complete the first round of the survey.

Appendix B provides full details of the survey analytic samples and tests for balance at baseline for clinic, quality of care, and household analytic samples. Overall, we find that randomization was effective at producing balanced, exchangeable comparison groups - all three analytic samples were, on average, balanced between treatment and control arms. We also find that AHME households at baseline were more likely to have improved measures of asset wealth, education, and household infrastructure compared to a representative sample of households from the same provinces.

3.5 Empirical Strategy

We estimate treatment effects on different sets of outcomes. The first are related to Structural and Process quality, a second set focuses on different measures of Financial sustainability, and a third group that explores the information asymmetries between providers and patients, and the role of altruism in explaining the observed changes in quality of care as a result of the AHME program. Since assignment to treatment was randomized and the experimental groups were balanced (see Appendix B), we estimate the effect of the program as the difference in means between treatment and control using linear models. When the unit of observation was the clinic (e.g. financial outcomes, structural quality, investments) we estimated the following model:

$$Y_i = \beta_0 + \beta_1 ITT_i + \varepsilon_i \tag{1}$$

where Y_i is the outcome variable for clinic i, ITT_i is a dummy variable indicating whether or not the clinic was assigned to the AHME group, β is the treatment effect and ε_i is the error term. Where necessary, we conducted additional sensitivity analyses to account for the effect of measurement error in self-reported financial data.

When the unit of observation was the client (e.g. exit interviews or SPs), the general model we estimated was:

$$Y_{ij} = \beta_0 + \beta_1 ITT_{ij} + X'_{ij}\beta + \eta_i + \varepsilon_{ij} \tag{2}$$

where Y_ij is the outcome variable for individual i and facility j, ITT_i is a dummy variable indicating whether or not the clinic was assigned to the AHME group, β is the treatment effect, X_{ij} is a relevant set of control variables, η_i are either clinic or SP and case fixed effects, and ε_{ij} is the error term. The standard errors were clustered at the clinic level to account for any intra-cluster correlation.

Indicators measured in monetary units were transformed using two monotonic, nonlinear transformations: Natural logarithm and hyperbolic arc sine (Burbidge et al. (1988)). We used natural logarithm for strictly positive values (revenues and expenses) and hyperbolic arc sine in the case of variables that naturally include zeros or negative values (such as profits or investments). For these models we report ITT estimates as semi-elasticities, and for the hyperbolic arc sine models we used the following expression:

$$\frac{ey}{dx} = \beta_1 \times \bar{T} \times \Big(\frac{\sqrt{\bar{y}^2 + 1}}{\bar{y}}\Big) \tag{3}$$

Where β_1 is the ITT parameter, $\bar{T}$ is the proportion of clinics in the treatment group, and $\bar{y}$ is the mean of the outcome of interest on its natural units (Bellemare and Wichman (2020)). Standard errors of elasticities were computed with the delta method using the *nlcom* command in Stata.

Statistical significance was evaluated using one-sided t-tests, where the null hypothesis is that the treatment effect is strictly negative ($H_0 : \beta_1 \leq 0$) for outcomes that were hypothesized to increase as a result of AHME (such as revenues or scale), and strictly positive ($H_0 : \beta_1 \geq 0$) for outcomes that were expected to decrease (such as unit costs).

We (asymptotically) controlled the family-wise error rate (FWER), or the probability of rejecting at least one true null hypothesis in a conceptually similar family of hypotheses under test at a pre-specified α level using the Romano & Wolf technique using the Stata command *rwolf* with 1000 bootstrap repetitions (Romano and Wolf, 2016). The clustering pattern was factored in for the bootstrap process and inference procedures where applicable (e.g. SP or household-level data).

3.6 Ethical Review

This study was granted clearance by the ethics committees within the AHME quantitative evaluation at Kenya Medical Research Institute (No. KEMRI/RES/7/3/1; NON-SSC PROTOCOL NO. 372) and the Human Subjects Committee for Innovations for Poverty Action IRB-USA (IPA IRB Protocol 1085). Similar to other SP studies with similar designs and embedded in an intervention (Kwan et al., 2019, 2018), we sought and received a waiver of provider informed consent to conduct the SP study from both ethics committees. All the SPs in this study were hired as field staff and participated in a 3-week training, 2-week pilot and were required to participate in refresher trainings throughout fieldwork in order to mitigate any potentially harmful events, such as unsafe injections, invasive tests, and consuming any medicines during encounters in the health sector.

4 Results

Central to AHME's supply-side strategy was a package of interventions designed to improve quality of care in private sector clinics across Kenya. Across these interventions, the AHME theory of change for quality improvement resided in the idea that greater access to high-quality care (generating a greater supply of patients) is necessary to translate solvent business practices and insurance coverage into better health outcomes for the poor. In this section, we report our results on the effects of AHME, first on management practices, then on quality of care and business outcomes.

4.1 Management Practices

First, we begin with our findings on management practices. The AHME intervention led to sizable and significant effects on the adoption and implementation in three out of four dimensions of management practices, as shown in table 6. The largest effects were concentrated in operations management (column 1), people management (column 4) and target setting (column 3). AHME clinics increased their compliance with Operations management in about 0.44 σ relative to the control group (p-value < 0.001). For people management, the estimated effect is 0.295 σ (p-value $= 0.016$), and 0.25 σ for target setting (p-value $= 0.055$). The effect on performance monitoring (column 2) and on the overall management index (column 5) were positive but estimated without precision at conventional levels.

4.2 Quality of Care

4.2.1 Structural Quality

We then turn to the effects on structural quality (table 7). We present the linear model results of two indices for structural quality compliance: interpersonal structural quality (16 of 33 indicators) and technical structural quality (17 of 33).[13] On average, clinics in the control group complied with 50% (SE 0.173) of interpersonal care indicators and 52% (SE 0.204) of technical care indicators; AHME treatment increased compliance by 8.4 (SE 0.025) and 7.7 (SE 0.028) percentage points, respectively. Graphing kernel densities of structural quality compliance conditional on treatment status suggests that clinics with the worst structural quality were most responsive to AHME treatment (see Appendix).

To explore the mechanisms by which the AHME program improved structural quality, we present the treatment effect on capital investments in table 7, columns 3-5. The AHME program increased investments in medical equipment by 86% (SE 0.49), but not investments in diagnostic or laboratory equipment. The AHME program also increased investments in amenities by 17% (SE 0.53), but the effect is not statistically significant (not shown).

For provider knowledge, we find knowledge to be high for diarrhea and malaria with AHME control clinic means at 91.9% and 98.4%, respectively (table 8). However, knowledge was particularly low for correctly managing the family planning case: 11% of the providers knew how to manage the case. This low knowledge level is likely due to how we define correct management for family planning. Four different actions need to occur during the interaction: the provider needs to ask a family planning history question; ask an obstetric history question; rule out pregnancy; and ask the client what method she prefers. When we consider a more lenient definition of case management for family planning where a provider knows at least one of the four actions, we find that knowledge is high. Providers who saw the

[13]In our sample of 199 clinics, Cronbach's alpha was 0.66 and 0.75 for structures to facilitate interpersonal and technical care, respectively. The social science literature considers an *alpha* greater than 0.7 to reflect a reliable instrument. Our instrument for structures to facilitate interpersonal care falls short of that rule of thumb. However, Cronbach's alpha for overall structural quality in our sample (0.83) is well above that.

family planning SPs knew to do at least one of the four actions for 94% of the interactions (not shown).

4.2.2 Process Quality

Next, we turn to whether improvements in interpersonal and technical structural quality and high provider knowledge translated into improvements in technical process quality. Beginning with correct actions, table 9a reports the effect of the AHME program on correct case management from SP data for diarrhea (column 1), family planning (column 2), asthma (column 3), malaria (column 4), and all cases pooled (column 5). We see that AHME treatment assignment decreased the likelihood of correct case management by 7.7 percentage points (p-value = 0.021; model 5) on average, compared to the control group (mean = 0.636, a relative reduction of 12%). When examining the different cases separately (models 1–4), we see that the coefficient on the AHME treatment indicator is significant only for the malaria scenario (coefficient = -0.092, p-value = 0.058). However, the coefficient on the AHME treatment is large and negative for each case, and pooling increases statistical power to detect the significant reduction in correct case management due to the AHME program. We depict the "know-do gap" (a well-documented phenomenon in the literature referring to the difference between provider knowledge and provider actions in practice) in Figure 2 (Das et al., 2015; Mohanan et al., 2015).

We also find that AHME treatment clinics are less likely to order any unnecessary lab tests (coefficient = -0.019 or 12.3% reduction, p-value = 0.423) (table 9b, model 5) and less likely to order any non-efficacious medicines (coefficient = -0.050 or 8.5% reduction, p-value = 0.139) (table 9c, model 5). However, the reductions in unnecessary actions in the pooled SP data are not statistically significant at α=0.10.

By examining panels (b) and (c) by case, we observe meaningful point estimates of reduction in unnecessary lab tests and unnecessary medicines of at least 10% magnitude, but it is not as precise as we would like. For example, we find a significant 51% reduction (coefficient = -0.070, p-value = 0.015) for the diarrhea SP case in AHME treatment clinics from the control group mean 0.138 (table 9b, model 1). The control group ordered low rates of any unnecessary tests for family planning (mean = 0.022, table 9b, model 2) and asthma (mean = 0.065) (table 9b, model 3) with negative effects from the AHME program assignment that were not significant at the $\alpha = 0.10$ level. The control group mean for any unnecessary lab tests for the malaria scenario was 0.289, with a 15% nonsignificant increase (p-value = 0.415) among the AHME treatment group (table 9b, model 4). Similarly for non-efficacious medicines, we find AHME treatment significantly reduced whether any non-efficacious medicines were ordered for the diarrhea case (coefficient = -0.106, p-value =

0.027) compared to AHME control group mean = 0.723 (table 9c, model 1) , but the AHME program assignment did not affect the other cases at the α=0.10 significance level (table 9c, models 2–4). Among all clinics, we find low rates of any unnecessary medicines for family planning (table 9c, model 2).

4.2.3 Health Care Outcomes

We complement our understanding of the AHME program's impact on the technical components of process quality by examining client perceptions of clinic amenities and health care outcomes, such as client satisfaction and household perceptions of clinic quality, from standardized patients (SPs), exit interviews with actual clients, and the household survey.

Table 10 shows the effects of AHME treatment on client perceptions of clinic amenities from the perspective of SPs (panel a) and real clients (panel b). First, we do not find evidence supporting the hypothesis that the AHME program improved amenities, based on our 5-item index (coefficient = 0.099, p-value = 0.182; model 1).[14] We do not find any significant program effects on whether SPs or clients found providers to spend sufficient amount of time with the nor whether they completely trusted the provider's medical treatment decision at the α=0.10 significance levels.

Further, we find no effect of AHME on household perceptions of clinic quality. Clients of AHME clinics at baseline did not rate those clinics higher than comparison clinics (p-value = 0.471; table 11, model 1). Clients in the AHME treatment group were also not more likely to rank the index clinic as high quality than clients in the control group (p-value = 0.302; table 11, model 2). That we observe that clients do not recognize the change in reduced correct care in the market demonstrates a market failure. We investigate this in the next part.

4.3 Provider Altruistic Preferences

Since clients and households did not recognize the observed changes due to AHME in the private sector for structural or process quality, we conjecture that this market failure could be due to asymmetries in medical information between private providers and clients. When patients are unable to discipline the markets by negotiating better care for themselves, it may be in the interest of purely profit maximizing clinics to reduce quality. Under such circumstances, high-quality health care could still prevail where regulation is strong and/or providers are willing to forgo profits for the well-being of their patient.

[14] This index consists of the following five items: whether the clinic was clean, whether the waiting time was appropriate, whether providers were courteous and respectful, whether clients had enough privacy, and whether operating hours were adequate.

Recognizing that preferences and attitudes shape behaviors, we first motivate our interpretation of provider preferences ("self-interest" as "profit driven") in the Kenyan private sector with SP narratives during fieldwork. We find anecdotal evidence for both altruistic and profit-driven behaviors. Regarding the former, we note that a small number of SPs portraying the poor case variant reported that concerned providers paid for necessary lab tests and medicines. Conversely, a small number of SPs reported instances in which the electricity did not work while the malaria microscopy test was being processed, rendering the ability to use a microscope impossible-although providers then proceeded to deliver a (false) positive test result to the SP.[15]

To further illustrate these different behaviors, we exhibit the following three narratives from SP fieldwork. (The reader can keep in mind that 100 Kenyan shillings is approximately 1 United States dollar.) These narratives describe how provider and patient exchanges reveal potential motives related to profit-driven behavior or altruism in the private sector:

- ***Female SP, seeking care for a sick child in absentia:*** *The doctor stepped away to talk to another doctor. They were communicating in Khamba, and they were negotiating how much to charge for the consultation fee. 600 [shillings]? Too much – 300 [shillings]. The doctor returned and asked me whether I was ok paying 300. I never shared that Khamba was my language, that I understood Khamba. He prescribed 2 types of syrups for 600 shillings and ORS for 20 [shillings]. They didn't have ORS, so I paid a total of 900 [shillings]. When they wrote the receipt, they didn't use carbon paper that was sitting there, but used scrap paper. At that clinic today, another SP was charged 200, and another was charged 400.*

- ***Male SP, presenting with malaria symptoms and a history of having taken artemether lumefantrine (AL):*** *She asked, have you taken any meds? AL and panadol. She said, well because you took meds – they are still in your system and RDT [rapid diagnostic test] is not going to work. So, if you are sick again, it must be typhoid. She wanted to do a blood draw. The test would cost 200 shillings and the meds for typhoid would be 2000. I said I didn't have money, so she asked, how much can you afford? I told her 300. "That's very little – just go back. If you have AL – go ahead, make sure you complete the dose." She didn't have a consultation fee, so I didn't need to pay anything.*

[15] A small number of SPs also reported instances in which they did not see malaria test results. We deduce that laboratory quality issues observed in the malaria case were motivated by profit considerations, rather than a lack of knowledge on the providers' parts regarding the diagnostic specificity of malaria diagnostic tests. If knowledge and skills were an issue, we would see false negatives because malaria parasites would not be picked up by a technician looking into a microscope. This is not the case, and instead we see high rates of false positives with onward treatment.

- ***Male SP, had difficulty breathing last night:*** *The lab tech was not in today, so the provider sent me to another clinic, just outside. He said the test was 500 shillings. As I was exiting, the receptionist was asked to accompany me. I don't have 500 shillings, I told them. The provider gave me – he insisted on it – 500 shillings and told me to go get that test.*

From these anecdotes, we observe that not all providers are completely profit oriented; not all providers are altruistic; and providers may be to some extent variable depending on the situation.

Following the literature, to estimate our social preference parameter, we perform checks on our data and confirm that each respondent's choices is consistent with individual utility maximization. We estimate the social preference parameter α in [0,1], which denotes the range from altruistic to fair-minded, following Andreoni and Miller (2002) and (Li et al., 2017). Individuals with $\alpha = 1$ ($\alpha = 0$) perfectly selfish (perfectly altruistic), as they put all weight on the payoff to self (other). Individuals with $\alpha = 0.5$ are fair-minded, as they put equal weight on payoffs to self and other. Thus $\alpha < 0.5$ reflects individuals who tend toward altruistic preferences, and $\alpha > 0.5$ reflects individuals who tend toward self-interested preferences.

To test our original hypothesis of whether the most profit-driven providers reduce correct care due to AHME, we examine the treatment effects by provider altruism on correct case management by including an interaction term between AHME treatment and each indicator representing whether the SP saw a provider falling into the 50% least altruistic, 25% least altruistic or 20% least altruistic group.[16] As we narrow in on the least altruistic (that is, most profit-driven) providers with our approach, we find that the most profit-driven providers at AHME clinics were the ones to significantly reduce correct care (see table 12).

More specifically, in table 12, model 1, we first include providers correctly identified as well as provider replacements for those providers who were not identified (matched on clinic and whether that provider sees patients for that service). Models 2–5 restrict the sample to SP visits conducted by correctly identified and matched providers from the modified dictator game. Model 1 does not include altruism parameters. Model 2 does not include altruism parameters and restricts the sample as described. Models 3, 4, and 5 include indicators for whether the provider seen by the SP is among the 50%, 25%, and 20% least altruistic, respectively of our continuous altruism parameter. Each model includes an interaction between AHME treatment (0, if clinic was in AHME control; 1, if clinic was in AHME treatment)

[16] We sequentially identify the bottom 50th percentile, bottom 25th percentile, and the bottom 20th percentile of providers on our altruism parameter, since there is no determined cutoff in the literature for identifying the providers who are the most self-interested.

and the indicator for whether the provider falls in the 50%, 25%, and 20% least altruistic group in order to identify heterogeneous treatment effects for the most for-profit providers versus those who are not.

Adjusting for this in our model demonstrates that the providers who are among the 80%, 75% and even 50% most altruistic do not contribute to the significant reduction in correct care due to the AHME program, since the coefficient on the AHME treatment indicator adjusting for the interactions is no longer significant at the 10% significance level (table 12, models 3, 4, 5). In parallel, what is striking is that the magnitude of the coefficients on the interactions get larger and increasingly significant (coefficient on AHME treatment and least altruistic (50%) in model 3: -0.159, p-value = 0.074; coefficient on AHME treatment and least altruistic (25%) in model 4: -0.189, p-value = 0.073; coefficient on AHME treatment and least altruistic (20%) in model 5: -0.227, p-value = 0.063, respectively).

Table 13 displays the heterogeneity of AHME treatment effects on total prices paid - using the inverse hyperbolic sine (IHST) transformation, based on Bellemare and Wichman (2020)—with interactions between AHME treatment and whether providers are among the 50%, 25%, or 20% least altruistic as described for our analysis on correct case management. We do not find evidence that the AHME program changed prices for patients at the 10% significance level (table 13, model 1). However, we find that SPs seeing the 20% least altruistic providers at AHME treatment clinics paid 121.2% more than the 20% least altruistic providers at AHME control clinics, and this was significant at the 1% significance level (semi-elasticity of least altruistic (20%) in AHME: 1.212, p-value = 0.003; table 13, model 5). For this same model, we see that AHME treatment clinics increased prices by 2% compared to the AHME control group mean, but this was not significant at the 10% significance level (semi-elasticity of AHME: 0.020, p-value = 0.693.

4.4 Business Outcomes

We next turn to the effects of AHME on several dimensions of business performance. The first group of outcomes total number of patients (scale), revenues, expenditures, profits, the probability that profits were positive, and the probability that the clinic was empaneled in the NHIF scheme. The second group of business outcomes were measures of efficiency (unit cost of clinical labor, unit cost of non-clinical labor, and unit cost of non-labor inputs).

We find a substantial increase in the total number of clients who demanded health services in the last full operational month as a result of AHME (table 14, model 1). Treatment clinics have, on average, 26% more clients (p-value = 0.056) relative to control clinics, which had 387 clients over the same period.

The intervention led to an increase in total revenues (table 14, model 2). On average, treatment clinics had 34% higher revenues (p-value = 0.042) relative to control clinics, which reported US$3,340 in the last full operational month. The effect on total expenses (table 14, model 3) is positive but small (3.8%) and is estimated without precision (p-value = 0.420). On average, expenses in control clinics were US$3,222 per month.

The treatment effects on profitability show that total profits increased in treatment clinics about 89% more than control clinics (p-value = 0.047) relative to an average of USD 170 among control clinics (table 14, model 4). The proportion of clinics that had positive profits increased as a result of AHME: compared to 71% of control clinics that had a positive profit in the last month, the treatment effect on the probability of having a positive profit is 8.2 percentage points (p-value = 0.097), as shown in model 5.

We also analyzed whether AHME was successful at empaneling clinics. We find that the intervention was successful in empaneling clinics with NHIF (table 14, model 6). Compared to an empanelment rate of 21% among control clinics, AHME increased the probability of clinic empanelment by 14.5 percentage points (p-value = 0.015), a relative increase of 70%.

NHIF empanelment (accreditation) comprised a big incentive for the private clinics under study. As noted in the midline qualitative interviews, "NHIF accreditation was increasingly common among providers, particularly in Kenya, by Round 3 of data collection. A number of providers were motivated to become accredited due to client demand, and thought that NHI(F) accreditation helped them better serve low-income populations. However, it was unclear whether increased client flow, even with accreditation, resulted in an increase in revenues and a more viable provider business model." The qualitative team also noted that "stronger outreach to potential patients to both connect them to the AHME clinics once enrolled in NHI(F) and educate them on NHI(F) benefits has the potential to increase client load in both countries, a perceived key benefit of AHME participation mentioned by many providers" (Suchman et al. (2017)).

The intervention led to sizable improvements in cost-efficiency. As shown in table 15, treatment clinics had 22% lower unit costs (p-value = 0.040) compared to an average of US$10.85 per client among control clinics. The unit cost breakdown shows that the treatment clinics engaged in cost-optimizing behavior for non-labor inputs (model 4), as treatment clinics had 26% lower costs than control clinics (p-value = 0.044), whose costs averaged US$7.82 per client. No significant or sizable changes were observed in unit labor costs (clinical or non-health staff).

5 Discussion

The AHME impact evaluation is the first study that examines truly comparable data among a clinic sample with a methodologically powerful combination of clinic surveys, provider vignettes, a modified dictator game designed to elicit provider preferences, the gold standard standardized patients, exit interviews with actual clients, and a household survey across the country of Kenya with high internal validity. We used the design and data to identify and estimate the causal impact of AHME on management outcomes, quality of care, and business outcomes, including financial viability, of private sector providers.

The first and foremost result of the evaluation is that AMHE was successfully implemented as planned. While about a quarter of the treatment clinics were disenfranchised, the rest had AHME interventions in place for at least 18 months before the endline evaluation. Hence, any null or negative results are most likely due to design issues as opposed to weak implementation problems.

In this study, we find that indeed management in health care is inherently different than other service industries previously studied. The AHME intervention successfully improved management outcomes and structural quality, but those improvements did not translate into better process quality or health care outcomes for outpatient services, despite high levels of provider knowledge and low rates of stockouts. Specifically, we find that providers in clinics assigned the AHME treatment had significantly lower levels of correct case management including minimally essential actions, and these reductions were accompanied by significantly higher prices. Since we found utilization, revenues, and prices increase due to the program without any changes to the number of clients in the waiting room and the amount of time a client spent with a provider, we believe these increases were absorbed by excess capacity.

Our findings suggest that distributive preferences among private providers, particularly altruism and self-interest, play a significant role in the provision of quality and affordable care. We find that AHME possibly changed the incentives of providers (that is, made them more profit-driven) without changing financial incentives. However, it could be possible that providers varied their preferences before participating in the AHME program and those that are more profit-driven responded to the program in a different way than other providers in the program. Both interpretations imply that business management interventions should be better aligned with process quality measures. The latter could alternatively imply that interventions should either be better tailored to provider preferences or conduct better signaling or screening measures to select providers according to their preferences. Understanding provider altruism can thus help identify strategies to better align future interventions with process quality measures and to strengthen data captured for monitoring purposes.

What might explain our results? The AHME program gave providers more understanding of and better control over quality, as well as an awareness of how expenditures on quality are related to profits. Profit-driven providers would likely have improved the quality of their services if the market sufficiently rewarded higher quality with increased demand (more patients) and greater willingness to pay (higher prices). However, the sensitivity of demand to quality depends on the ability of patients to recognize quality improvements, value those improvements, and be willing to pay for improvement. One of the most studied market failures in health care since Arrow (1963) is the asymmetric information between patients and providers regarding quality of care; namely, patients are unable to judge quality of care. If patients do not recognize adjustments in quality and are not willing to pay for them, then there is no gain in profit from increasing quality. In this case it would be more profitable to lower quality and thereby reduce costs without affecting revenues. Indeed, we found substantial evidence to support the notion that patient demand was insensitive to the drop in process quality based on measures of satisfaction with care from the SP surveys, household surveys, and exit surveys.

These results imply that interventions that simply train providers to improve process knowledge, structural quality, and management practices are not likely to succeed in improving technical process quality. This is likely to be true regardless of provider preferences for profits versus altruism. Simply scaling up the current combination of interventions in AHME is unlikely to work without more attention to process quality.

One such way to adjust the current design would be to include interventions that emphasize accountability, such as monitoring, effective regulation, and financial and non-financial incentives for providers to deliver better process quality. For example, Government regulators or an external monitoring body could assess and make public clinical quality assessments. If such assessments influence demand for and sensitivity to clinic's quality of care, market could reward clinics for higher quality and punish them for lower quality. Alternatively, NHIF payments could be conditioned on quality in a pay for performance model.

Our results also imply that interventions need to account for how variation in provider preferences for profits versus altruism affects how providers respond to the incentives and opportunities created by the interventions. Increasing accountability, for example, is especially important for providers who are less altruistic and more profit driven. Another avenue to is consider is changing norms, such as an emphasis on fair-mindedness or altruistic preferences. We conclude that AHME did make significant strides toward improving health equity, but there remain substantial inequities. Both supply-side and demand-side engagement in NHIF were crucial to these achievements. We found that AHME clinics reached out to more poor and disadvantaged households within their catchment areas, but that outreach had limited effects in terms of patient composition. An important caveat to consider is that we were only able to analyze how AHME affected access to care by the poor among those who lived in the catchment areas of the clinics in the evaluation sample and not how the choice of clinics that would be eligible for the AHME program affected access. One way to greatly increase access to care by the poor is to offer the AHME program to clinics that are located in poor areas. The evaluation was not designed to address this question. AHME made progress in the expansion of NHIF both at the patient and clinic level. While AHME expanded clinic NHIF empanelment, empanelment only reached about one-third of the clinics. Consequently, many of the newly insurance patients may not have had access to facilities where they could use that insurance. While the share of patients with health insurance rose in treatment areas relative to control areas, we found no differential expansion among the poor relative to the nonpoor, meaning that AHME was not able to dramatically expand health insurance among the poor. We found evidence that the poor sought care less and received lower quality of care when they did, highlighting the critical role of health insurance in alleviating many of the financial barriers facing the poor. While low-cost and high-quality insurance such as NHIF is important for the financial and health protection of the poor, take up and use of health insurance are difficult behaviors to change. Despite AHME's achievements, even larger gains in health insurance coverage of the poor and clinic NHIF empanelment in poorer areas is critical for substantially improving health equality.

Finally, the AHME program was most successful in improving the financial viability of private for-profit clinics. Clinics appeared to absorb the management practices being promoted and were able to expand scale (patient load), increase revenue, and become more efficient in the delivery of care. They improved efficiency both through economies of scale and through more efficient management techniques such as negotiating lower prices for laboratory supplies. However, some of these efficiency gains may have been achieved by lowering quality. Business practice interventions need to be better integrated with quality interventions to avoid this unintended outcome.

Bibliography

AHME Program (2016). Ahme business improvement plan example.

Anderson, S. J., C. M. Lazicky, and B. H. Zia (2019). *Measuring the Unmeasured: Combining Technology and Behavioral Insights to Improve Measurement of Business Outcomes*. The World Bank.

Andreoni, J. and J. Miller (2002). Giving according to garp: An experimental test of the consistency of preferences for altruism. *Econometrica 70*(2), 737–753.

Balakrishnan, U., J. Haushofer, and P. Jakiela (2020). How soon is now? evidence of present bias from convex time budget experiments. *Experimental Economics 23*(2), 294–321.

Barnes, J., B. O'Hanlon, F. Feeley, M. Kimberly, G. Nelson, and D. Caytie (2010). *Private health sector assessment in Kenya*. The World Bank.

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

Bloom, N., B. Eifert, A. Mahajan, D. McKenzie, and J. Roberts (2013). Does management matter? evidence from india. *The Quarterly Journal of Economics 128*(1), 1–51.

Bloom, N., R. Lemos, R. Sadun, and J. Van Reenen (2020). Healthy business? managerial education and management in health care. *Review of Economics and Statistics 102*(3), 506–517.

Bloom, N., A. Mahajan, D. McKenzie, and J. Roberts (2018). *Do management interventions last? evidence from India*. The World Bank.

Bloom, N., R. Sadun, and J. Van Reenen (2014). Does management matter in healthcare. *Boston, MA: Center for Economic Performance and Harvard Business School*.

Bruhn, M., D. Karlan, and A. Schoar (2018). The impact of consulting services on small and medium enterprises: Evidence from a randomized trial in mexico. *Journal of Political Economy 126*(2), 635–687.

Currie, J., W. Lin, and W. Zhang (2011). Patient knowledge and antibiotic abuse: Evidence from an audit study in china. *Journal of health economics 30*(5), 933–949.

Daniels, B., A. Dolinger, G. Bedoya, K. Rogo, A. Goicoechea, J. Coarasa, F. Wafula, N. Mwaura, R. Kimeu, and J. Das (2017). Use of standardised patients to assess quality of healthcare in nairobi, kenya: a pilot, cross-sectional study with international comparisons. *BMJ global health 2*(2), e000333.

Das, J. and J. Hammer (2014). Quality of primary care in low-income countries: facts and economics. *Annu. Rev. Econ. 6*(1), 525–553.

Das, J., A. Holla, V. Das, M. Mohanan, D. Tabak, and B. Chan (2012). In urban and rural india, a standardized patient study showed low levels of provider training and huge quality gaps. *Health affairs 31*(12), 2774–2784.

Das, J., A. Kwan, B. Daniels, S. Satyanarayana, R. Subbaraman, S. Bergkvist, R. K. Das, V. Das, and M. Pai (2015). Use of standardised patients to assess quality of tuberculosis care: a pilot, cross-sectional study. *The Lancet infectious diseases 15*(11), 1305–1313.

Donabedian, A. (1966). Evaluating the quality of medical care. *The Milbank memorial fund quarterly 44*(3), 166–206.

Donabedian, A. (1978). The quality of medical care. *Science 200*(4344), 856–864.

Donabedian, A. (1988). The quality of care: how can it be assessed? *Jama 260*(12), 1743–1748.

Dunsch, F. A., D. K. Evans, E. Eze-Ajoku, and M. Macis (2017). Management, supervision, and health care: A field experiment. Technical report, National Bureau of Economic Research.

Evans, T., M. Whitehead, and F. Diderichsen (2001). *Challenging inequities in health: from ethics to action*. Oxford University Press.

Fafchamps, M., D. McKenzie, S. Quinn, and C. Woodruff (2012). Using pda consistency checks to increase the precision of profits and sales measurement in panels. *Journal of Development Economics 98*(1), 51–57.

Fisman, R., P. Jakiela, and S. Kariv (2017). Distributional preferences and political behavior. *Journal of Public Economics 155*, 1–10.

Fitzpatrick, A. and K. Tumlinson (2017). Strategies for optimal implementation of simulated clients for measuring quality of care in low-and middle-income countries. *Global Health: Science and Practice 5*(1), 108–114.

Gertler, P., C. Boone, D. Contreras-Loya, R. Cuckovich, J. Gruber, A. Kwan, and N. Perales (2020). African health markets for equity quantitative evaluation final report, 2013 – 2019.

Gertler, P. J. and D. M. Waldman (1992). Quality-adjusted cost functions and policy evaluation in the nursing home industry. *Journal of Political Economy 100*(6), 1232–1256.

Hurley, J. (2000). An overview of the normative economics of the health sector. In *Handbook of health economics*, Volume 1, pp. 55–118. Elsevier.

Jakiela, P. (2013). Equity vs. efficiency vs. self-interest: on the use of dictator games to measure distributional preferences. *Experimental Economics 16*(2), 208–221.

Karlan, D. and M. Valdivia (2011). Teaching entrepreneurship: Impact of business training on microfinance clients and institutions. *Review of Economics and statistics 93*(2), 510–527.

Kenya Ministry of Health (2010a). *National Family Planning Guidelines for Service Providers*. Kenya Ministry of Health, Division of Reproductive Health.

Kenya Ministry of Health (2010b). *National Guidelines for Diagnosis, Treatment and Prevention of Malaria in Kenya*. Kenya Ministry of Health.

Kenya Ministry of Health (2014). *Policy Guidelines for Management of Diarrhoea in Children below Five Years in Kenya*. Kenya Ministry of Health.

Kenya Ministry of Public Health and Sanitation (2010). *Guidelines for Asthma Management in Kenya*. Kenya Ministry of Health.

Kenya President's Malaria Initiative (2018). *Kenya Malaria Operational Plan FY 2018*. President's Malaria Initiative.

King, J. J., J. Das, A. Kwan, B. Daniels, T. Powell-Jackson, C. Makungu, and C. Goodman (2019). How to do (or not to do)... using the standardized patient method to measure clinical quality of care in lmic health facilities. *Health policy and planning 34*(8), 625–634.

Kwan, A., B. Daniels, S. Bergkvist, V. Das, M. Pai, and J. Das (2019). Use of standardised patients for healthcare quality research in low-and middle-income countries. *BMJ global health 4*(5), e001669.

Kwan, A., B. Daniels, V. Saria, S. Satyanarayana, R. Subbaraman, A. McDowell, S. Bergkvist, R. K. Das, V. Das, J. Das, et al. (2018). Variations in the quality of tuberculosis care in urban india: A cross-sectional, standardized patient study in two cities. *PLoS medicine 15*(9), e1002653.

Li, J., W. H. Dow, and S. Kariv (2017). Social preferences of future physicians. *Proceedings* *of the National Academy of Sciences 114*(48), E10291–E10300.

McKenzie, D. (2020). Small business training to improve management practices in developing countries.

McKenzie, D. and C. Woodruff (2014). What are we learning from business training and entrepreneurship evaluations around the developing world? *The World Bank Research Observer 29*(1), 48–82.

Mohanan, M., M. Vera-Hernández, V. Das, S. Giardili, J. D. Goldhaber-Fiebert, T. L. Rabin, S. S. Raj, J. I. Schwartz, and A. Seth (2015). The know-do gap in quality of health care for childhood diarrhea and pneumonia in rural india. *JAMA pediatrics 169*(4), 349–357.

NMCP, KNBS, and ICF. Kenya malaria indicator survey 2015.

Peabody, J. W., J. Luck, P. Glassman, T. R. Dresselhaus, and M. Lee (2000). Comparison of vignettes, standardized patients, and chart abstraction: a prospective validation study of 3 methods for measuring quality. *Jama 283*(13), 1715–1722.

PharmAccess (2017a). Safecare basic healthcare standards advanced assessment report.

PharmAccess (2017b). Safecare basic healthcare standards basic assessment report.

PharmAccess (2017c). Safecare basic healthcare standards quality improvement plan.

PharmAccess (2017d). Safecare basic healthcare standards quality improvement plan.

PharmAccess Group (2017). Safecare overview.

Population Services Kenya (2016). Ahme business assessment example.

Romano, J. P. and M. Wolf (2016). Efficient computation of adjusted p-values for resampling-based stepdown multiple testing. *Statistics & Probability Letters 113*, 38–40.

Suchman, L., D. Montagu, and A. Seefeld (2017). African health markets for equity qualitative evaluation comprehensive report, 2013–2017. Institute for Global Health Sciences, University of California, San Francisco.

Sylvia, S., Y. Shi, H. Xue, X. Tian, H. Wang, Q. Liu, A. Medina, and S. Rozelle (2014). Survey using incognito standardized patients shows poor quality care in china's rural clinics. *Health policy and planning 30*(3), 322–333.

Walker, F. A. (1887). The source of business profits. *The Quarterly Journal of Economics 1*(3), 265–288.

Figures and Tables

Figure 1: Map of AHME Evaluation Sample (N=232 Clinics)

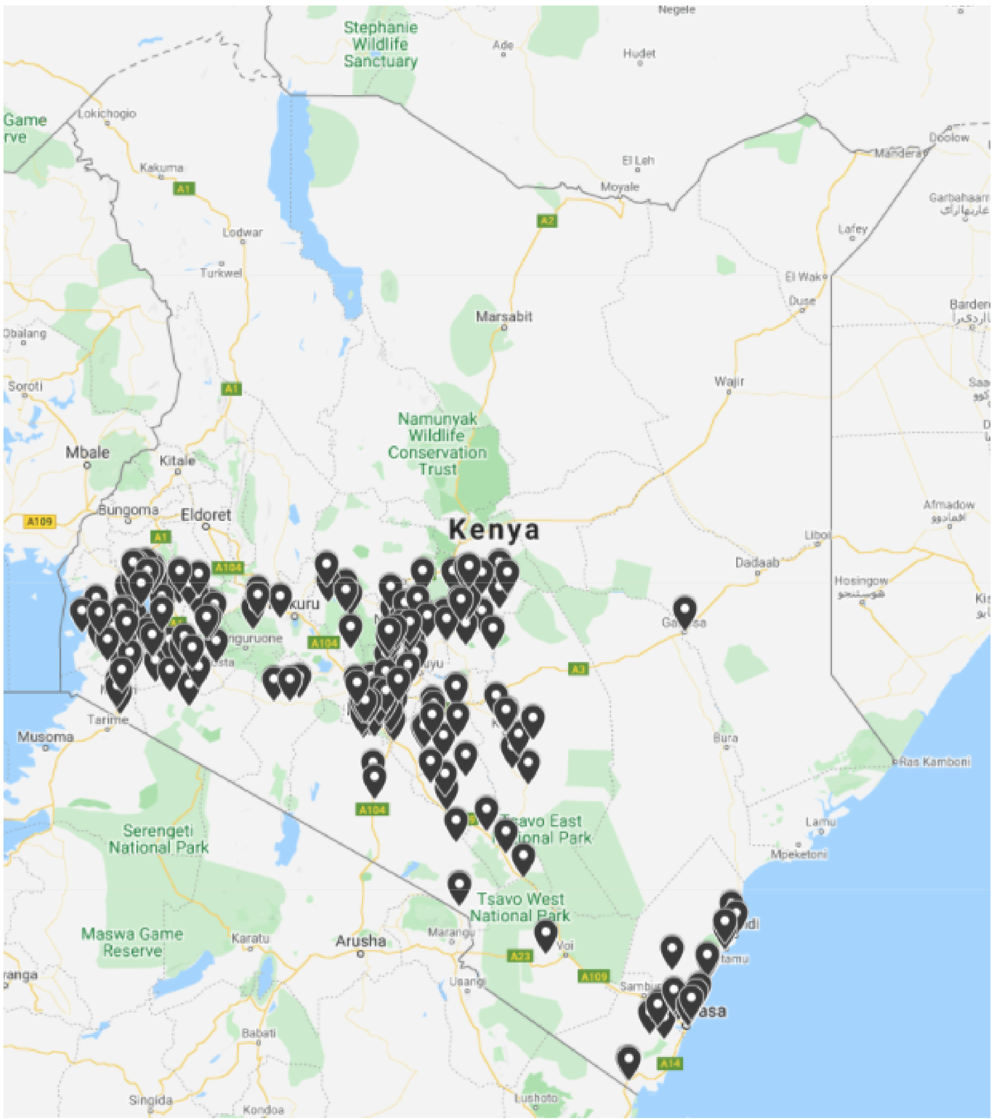
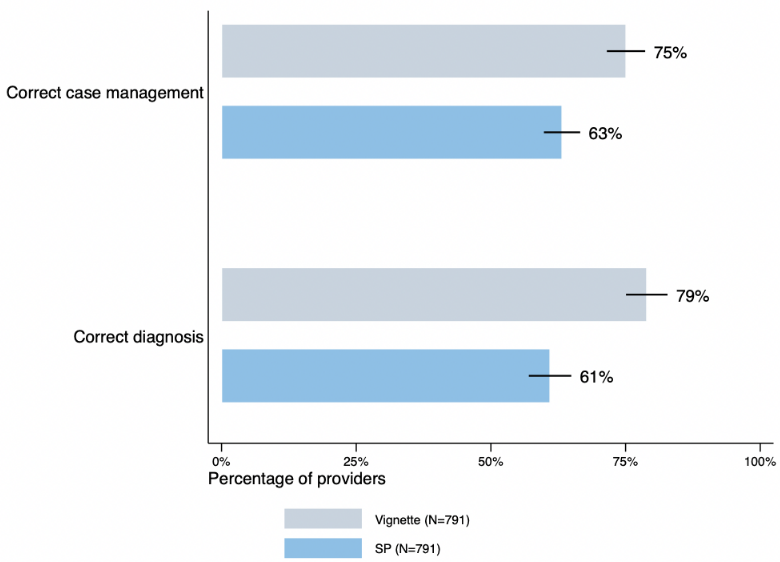


Figure 2: Differences in Provider Knowledge and Practice



Note: The figure shows the means and 95% confidence intervals from vignette (grey) and standardized patient (SP) (blue) data. The bar graphs depict averages for correct case management (top two bars) and correct diagnosis (bottom two bars) with 95% confidence intervals, comparing provider-matched vignette data and SP data. The percentage at the end of the bar is the percentage of providers for all SP visits who correctly managed the case or correctly diagnosed the case in the vignette or in practice.

Table 1: Conceptual Framework for Quality of Care Adapted from Donabedian (1966, 1978, 1988)

	Structure Attributes of material resources, human resources, and organizational structure in the setting where care occurs	**Process** Patients seeking and receiving care, as well as providers diagnosing and treating	**Outcomes** Health status, as well as patient knowledge, behaviors, and satisfaction
Interpersonal Exchange through which patients and providers share information and preferences	Dimension 1	Dimension 2	Dimension 3
Technical Recommendation of the appropriate strategies of care and implementation of those strategies	Dimension 4	Dimension 5	Dimension 6

Table 2: Standardized Patient (SP) and Vignette Case Scenarios for Assessing Process Quality

Case	SP	Vignette	Case description	SP opening statement	SP experiments
Childhood Diarrhea (Watery)	Yes	Yes	A 28-year-old mother comes to the clinic; her 1.5-year-old child is at home sick with diarrhea.	"My child has been having diarrhea."	Pre-demanding vs. Demanding albendazole vs. Demanding amoxicillin
Family Planning	Yes	Yes	A 27-year-old married female seeks family planning advice since she does not want to have children for another 2 years.	"I do not wish to have any other children for 2 years, and I would like to know what methods are possible."	Married vs. 25-year-old unmarried female
Acute Asthma	Yes	No	A 24/25-year-old female/male presents with great difficulty breathing the previous night.	"Doctor, last night I had a lot of difficulty with breathing."	Standard case vs. Cannot afford more than KSH 300 ("Poor")
Acute Malaria	Yes	Yes	A 28-year-old female or male who has recently traveled to a malaria-endemic area presents with fever and headache and thinks s/he has malaria.	"Doctor, I think I have malaria."	Standard case vs. Cannot afford more than KSH 300 ("Poor")

Table 3: Process Quality Main Outcomes

Case	Correct Case Management	Lab Tests (Valid or Unnecessary)	Medicines (Efficacious or Non-Efficacious)
Childhood diarrhea	Gave or advised on ORS, or Referred or asked to return	**Valid:** Stool test **Unnecessary:** All other tests	**Efficacious:** ORS, zinc **Non-efficacious, harmful:** Amoxicillin **Non-efficacious, harmless:** Albendazole **Non-efficacious:** All other meds
Family planning	Asked FP method history, and Asked obstetric history, and Ruled out pregnancy, and Asked preferred FP method	**Valid:** Pregnancy test **Unnecessary:** All other tests	**Efficacious:** Contraceptive pills **Non-efficacious:** All other meds
Asthma	Treated with inhaler or bronchodilator (e.g. salbutamol, cetirizine, prednisolone)	**Unnecessary:** All tests	**Efficacious:** Salbutamol, cetirizine, prednisolone, any other inhaler or bronchodilator **Non-efficacious:** Franol, any other meds
Malaria	Ordered malaria rapid diagnostic test (mRDT) or malaria microscopy	**Valid:** mRDT, malaria microscopy, blood count, brucellosis **Unnecessary:** All other tests	**Efficacious:** Artemether lumefantrine (AL), paracetamol **Non-efficacious:** All other meds

Note: FP refers to family planning; ORS refers to oral rehydration salts; SP refers to standardized patient.

Table 4: Utilization Indicators

Indicator	Description	Examples
Total number of clients	Total number of health visits in the last full operational month	All diagnostic, curative and preventive health visits, for all age groups
General outpatient clients	Total number of curative visits, including adults and children, over the last full operational month	Health visits with diagnostic and treatment of illnesses, e.g. gastroenteritis, respiratory syndromes, wounds stitching, etc.
Maternal and child health clients	Total number of health visits for mothers and their children in the last full operational month	Antenatal Care, Postnatal care, Family Planning, Immunizations, and Well-baby visits
Inpatient clients	Total number of patient admissions for observation in the last full operational month	Patients who were admitted for observation for more than 8 hours, and used some sort of bed facility

Table 5: AHME Revenue and Cost Categories with Descriptions and Examples)

Financial outcome	Description	Example
Revenues	Monthly-equivalent total revenues in the last full operational month	1) Revenues from out-of-pocket clients; 2) revenues from NHIF claims, either fee-for-service or capitation; 3) revenues from commercial insurers claims
Expenses	Monthly-equivalent total expenses in the last full operational month	Includes all variable and fixed costs incurred in the last full operational month: Payroll, utilities, transportation and communication, lab tests, medicines, rent of building, rent of equipment, maintenance costs, outsourcing costs (e.g. waste disposal), loans, fees, and taxes.
Profits	Difference between total revenues and total expenses	May not match actual "books" profit since we attempted to measure economic profit rather than financial profit (e.g. extraordinary costs were omitted, such as corrective maintenance)
Positive profit	Binary indicator with value = 1 if the calculated profit was equal or larger than zero	
Labor costs	Total staff expenses in the last full operational month	Salaries, contractual jobs expenses, per diem, financial incentives for clinical and support staff. In total and per staff cadre (doctors, nurses, health support staff, administrative staff, support staff)
Non-labor costs	Total variable and fixed non-labor costs in the last full operational month	Utilities, rent of building, rent of equipment, transportation costs, health inputs (medicines and laboratory inputs), cleaning and other outsourced services, waste disposal, security, permits, fees, taxes, and loans.
Investments	One-time purchases or acquisitions usually related to equipment, not including maintenance of replacement costs	Medical equipment, Information Technologies equipment, Diagnostic and Laboratory Equipment, Amenities (e.g. furniture, lockers, air conditioning)

Table 6: Average Treatment Effects on Management Practices

	(1) Operations management index	(2) Performance monitoring index	(3) Target setting index	(4) People management index	(5) Overall management index
AHME treatment					
Coefficient	0.444	0.045	0.260	0.295	0.106
Standard error	(0.143)	(0.145)	(0.162)	(0.137)	(0.146)
p-value	[0.001]	[0.378]	[0.055]	[0.016]	[0.234]
Mean of control	0.000	0.000	0.000	0.000	0.000
Observationss	199	199	198	199	199

Note: Standard errors in parentheses; one-sided p-values are in brackets ($H_0 : \beta \leq 0$). Effect sizes are standard deviations with respect to the mean of the control group.

Table 7: Average Treatment Effects on Structural Quality and Investments

	(1) Interpersonal structural quality	(2) Technical structural quality	(3) Total investment in diagnostic equipment	(4) Total investment in laboratory equipment	(5) Total investment in medical equipment
AHME treatment					
Coefficient	0.084	0.077	-0.195	-0.225	0.866
Standard error	(0.025)	(0.028)	(0.333)	(0.448)	(0.494)
p-value	[0.000]	[0.003]	[0.720]	[0.692]	[0.041]
Adjusted p-value	<0.003>	<0.012>	<0.861>	<0.861>	<0.092>
Mean control group	0.50	0.52	76	689	308
Observations	187	187	187	187	187

Note: Standard errors in parentheses; one-sided p-values are in brackets ($H_0 : \beta \leq 0$); adjusted p-values for multiple hypothesis testing (Romano & Wolf) in angled brackets. Models 1 and 2 measure effects on compliance, which takes a value of 0 to 1 and represents the share of associated indicators met in the clinic. Models 3-5 measure the effect on investments (between 2018 and 2019) in the following categories: diagnostic equipment (microscope, glucometer, sphygmomanometer, and ophtalmoscope), laboratory equipment (biochemistry analyzer, ELISA reader, centrifuge, hemogram machine, ultrasound and X-ray machines), medical equipment (beds, examination tables, negatoscope, oxygen machine, refrigerator, steriliser, ventilator machine). The inverse hyperbolic sine transformation (IHST) was utilized to deal with zeros, and its functional form is $arcsinh(x) = ln[x + \sqrt{(}x2 + 1)]$.

Table 8: Average Treatment Effects on Knowledge of Correct Case Management

	(1) Diarrhea	(2) Family Planning	(3) Malaria
AHME treatment			
Coefficient	-0.016	0.039	-0.008
Standard Error	(0.032)	(0.044)	(0.017)
p-value	[0.628]	[0.374]	[0.624]
Mean control group	0.919	0.107	0.984
Observations	288	285	287

Note: Standard errors in parentheses; two-sided p-values are in brackets. The table shows multivariate regressions using provider survey data. Knowledge of correct case management is a 0-1 binary measure. For diarrhea, knowledge of correct management = 1 is if the provider mentioned giving or advising on oral rehydration salts or gave a referral or asked the vignette scenario to return; 0 otherwise. For family planning (all 4 components), vignette data were coded as knowledge of correct management = 1 if the provider performed all four of the following actions: asked any family planning history question, asked any obstetric history question, ruled out pregnancy (by asking or offering a test), and asking the client her preferred family planning method; 0 otherwise. For malaria, vignette data were coded as knowledge of correct management = 1 if the provider mentioned ordering a malaria rapid diagnostic test (RDT) or a malaria microscopy test; 0 otherwise.

Table 9: Average Treatment Effects on Process Quality: SP data

	(1) Diarrhea	(2) Family planning	(3) Asthma	(4) Malaria	(5) Pooled
(a) Correct Case Management					
AHME treatment					
Coefficient	-0.069	-0.084	-0.069	-0.092	-0.077
Standard Error	(0.048)	(0.060)	(0.069)	(0.049)	(0.033)
p-value	[0.155]	[0.165]	[0.324]	[0.058]	[0.021]
Mean control group	0.686	0.264	0.5	0.85	0.636
(b) Any Unnecessary Lab Tests					
AHME treatment					
Coefficient	-0.070	-0.009	-0.028	0.044	-0.019
Standard Error	(0.029)	(0.021)	(0.031)	(0.054)	(0.024)
p-value	[0.015]	[0.656]	[0.366]	[0.415]	[0.423]
Mean control group	0.138	0.022	0.065	0.289	0.154
(c) Any Unnecessary Medicines					
AHME treatment					
Coefficient	-0.106	-0.043	-0.058	0.005	-0.05
Standard Error	(0.048)	(0.039)	(0.072)	(0.048)	(0.033)
p-value	[0.027]	[0.269]	[0.420]	[0.914]	[0.139]
Mean control group	0.723	0.088	0.554	0.717	0.586
Observations	400	199	204	392	1195

Note: The table shows multivariate regressions using standardized patient (SP) data. Standard errors are in parentheses for models 1–4. Robust standard errors (in parentheses) are clustered at the clinic level for pooled model 5. Two-sided *p*-values are in brackets. All models contain SP fixed effects and contain covariates for the 0-1 AHME treatment indicator and a 0-1 indicator for each SP experiment (demanding, unmarried, poor). Model 5 contains case fixed effects. Correct case management is a 0-1 binary measure constructed specific to each scenario.

Table 10: Average Treatment Effects on Clients' Perception and Satisfaction With Care

	(1) 5-item Amenities index	(2) Providers spent sufficient time (=1)	(3) Completely trust provider's medical treatment decision (=1)
(a) Standardized Patients			
AHME treatment			
Coefficient	0.099	0.022	-0.023
Standard Error	(0.074)	(0.037)	(0.031)
p-value	[0.182]	[0.561]	[0.450]
Adjusted p-value	<0.2238>	<0.4815>	<0.7672>
Mean control group	0.000	0.498	0.782
Observations	1086	1086	1076
(b) Exit Interviews			
AHME treatment			
Coefficient	0.075	-0.033	0.032
Standard Error	(0.074)	(0.026)	(0.023)
p-value	[0.306]	[0.204]	[0.163]
Adjusted p-value	<0.2847>	<0.0899>	<0.2088>
Mean control group	0.000	0.544	0.888
Observations	1539	1531	1528

Note: Standard errors in parentheses; one-sided p-values are in brackets ($H_0 : \beta \leq 0$); adjusted p-values for multiple hypothesis testing (Romano & Wolf) in angled brackets. The table shows multivariate regressions for the same set of outcomes on client's perception and satisfaction with care using standardized patient (SP) and exit interviews (EI) data in the top and bottom panels, respectively. Robust standard errors (in parentheses) are clustered at the clinic level. All models in the upper panel contain SP and case fixed effects and control for the 0-1 AHME treatment indicator and a 0-1 indicator for each SP experiment (demanding, unmarried, poor). The dependent variable for Model 1 is a five-item index standardized to ($\mu_{AHMETreatment=0} = 0, \sigma = 1$) for the five perceptions of clinic amenities: (i) whether the clinic was clean (=1), (ii) whether the waiting time was appropriate (=1), (iii) whether provider seen was courteous and respectful (=1), (iv) whether client had enough privacy (=1), and (v) whether operating hours were adequate (=1). Models 2 and 3 relate to SP or client perceptions about the care received: whether providers spent sufficient time with the client (=1) and whether the client completely trusted the provider's treatment decision (=1), respectively.

Table 11: Average Treatment Effects on Household Perceptions of the Index Clinic

	(1) Index clinic quality rating 1 (low)–10 (high) (z-score)	(2) Index clinic ranked as high quality for visit (z-score)
AHME treatment		
Coefficient	0.039	0.05
Standard Error	(0.054)	(0.048)
p-value	[0.471]	[0.302]
Mean control group	0.000	0.000
Observations	2637	2637
Households	1139	1139
Clinics	192	192

Note: The table shows multivariate regressions using household survey data. Standard errors in parentheses, clustered at level of index clinic; two-sided p-values are in brackets. Clients were asked to rank the quality of services at the index clinic compared to other clinics they had visited. The dependent variable in model 1 is the numerical rating of the index clinic, transformed into a z-score with mean control group zero. The dependent variable in model 2 is an indicator for if the index clinic was ranked in the 1st–50th percentile among the respondent's choice set, where a higher quality clinic is ranked lower. Households gave their opinion on three types of visits: child curative, child preventative, and prenatal care. These models include responses about all three types of visits. We control for the type of visit, where prenatal care is the reference category (regression coefficients not shown).

Table 12: AHME Treatment Effects by Provider Altruism on Correct Case Management

	(1)	(2)	(3)	(4)	(5)
	Correct case management				
AHME treatment					
Coefficient	-0.066	-0.049	0.024	-0.01	-0.011
Standard Error	(0.033)	(0.045)	(0.061)	(0.050)	(0.048)
p-value	[0.048]	[0.278]	[0.690]	[0.850]	[0.822]
AHME treatment × Least Altruistic (50%)					
Coefficient			-0.159		
Standard Error			(0.088)		
p-value			[0.074]		
AHME treatment × Least Altruistic (25%)					
Coefficient				-0.189	
Standard Error				(0.105)	
p-value				[0.073]	
AHME treatment × Least Altruistic (20%)					
Coefficient					-0.227
Standard Error					(0.121)
p-value					[0.063]
Mean control group	0.690	0.663	0.663	0.663	0.663
Observations	1086	674	674	674	674

Note: The table shows multivariate ordinary least square (OLS) regressions using standardized patient (SP) data and provider survey data from a modified dictator game where each observation is one SP-provider interaction. Robust standard errors clustered at the clinic level are in parentheses; two-sided p-values are in brackets. All models include covariates for AHME treatment (0 if clinic was in AHME control; 1 if clinic was in AHME treatment) and for each SP experiment (demanding, unmarried, poor); and control for SP and case fixed effects. The outcome correct case management is a binary indicator for whether the SP visit was correctly managed. Model 1 includes the full sample with providers correctly identified and replacements for those providers who were not identified (same-clinic and same-service). Models 2–5 restrict the sample to SP visits conducted by correctly identified and matched providers from the provider survey's dictator game. Model 1 does not include altruism parameters. Model 2 does not include altruism parameters and restricts the sample as described. Models 3, 4, and 5 include indicators for whether the provider seen by the SP is 50%, 25%, 20% least altruistic, respectively, based on our continuous altruism parameter α : Models 3,4,5 includes a binary variable to indicate whether the provider falls in the 50%, 25%, 20% least altruistic group and interactions between AHME treatment (0-1) and the binary variable for whether the provider falls in the 50%, 25%, 20% least altruistic group. Altruistic indicators are the bottom percentage of $-\alpha$: the bottom 50% is the 50% least altruistic (that is, most self-interested), ..., and the bottom 20% is the 20% least altruistic. Individuals with $\alpha = 1$ ($\alpha = 0$) are perfectly selfish (perfectly altruistic), as they put all weight on the payoff to self (other). Individuals with $\alpha = 0.5$ are fair minded, as they put equal weight on payoffs to self and other.

Table 13: AHME Treatment Effects by Provider Altruism on Prices

	(1)	(2)	(3)	(4)	(5)
	Price of visit in Kenyan Shillings, IHST)				
AHME treatment					
Coefficient	0.103	0.347	0.053	0.091	0.104
Standard Error	(0.177)	(0.255)	(0.339)	(0.273)	(0.263)
p-value	[0.560]	[0.174]	[0.876]	[0.738]	[0.694]
AHME treatment × Least Altruistic (50%)					
Coefficient			0.712		
Standard Error			(0.563)		
p-value			[0.207]		
AHME treatment × Least Altruistic (25%)					
Coefficient				1.509	
Standard Error				(0.737)	
p-value				[0.042]	
AHME treatment × Least Altruistic (20%)					
Coefficient					2.233
Standard Error					(0.822)
p-value					[0.007]
Semi-elasticity of AHME					0.020
p-value AHME					0.693
Semi-elasticity of Least Altruistic (20%) in AHME					1.212
p-value Least Altruistic (20%) in AHME					0.003
Mean control group	5.367	5.135	5.135	5.135	5.135
Observations	684	401	401	401	401

Note: Robust standard errors clustered at the clinic level are in parentheses; two-sided p-values are in brackets. These data exclude SP visits for childhood diarrhea pre-demanding (since price is captured once post-demanding) and asthma poor and malaria poor (since price was capped by design at KSH 300). All models include covariates for each SP experiment (unmarried), and control for SP and case fixed effects. The outcome is the total amount the SP paid (in Kenyan shillings, KSH) with inverse hyperbolic sine transformation (IHST). Model 1 includes the full sample with providers correctly identified and replacements for those providers who were not identified (same-clinic and same-service). Models 2–5 restrict the sample to SP visits conducted by correctly identified and matched providers from the provider survey's dictator game. Model 1 does not include altruism parameters. Model 2 does not include altruism parameters and restricts the sample as described. Models 3, 4, and 5 include indicators for whether the provider seen by the SP is 50%, 25%, 20% least altruistic, respectively, based on our continuous altruism parameter α : Models 3,4,5 includes a binary variable to indicate whether the provider falls in the 50%, 25%, 20% least altruistic group and interactions between AHME treatment (0-1) and the binary variable for whether the provider falls in the 50%, 25%, 20% least altruistic group. Altruistic indicators are the bottom percentage of $-\alpha$: the bottom 50% is the 50% least altruistic (that is, most self-interested), ..., and the bottom 20% is the 20% least altruistic. Individuals with $\alpha = 1$ ($\alpha = 0$) are perfectly selfish (perfectly altruistic), as they put all weight on the payoff to self (other). Individuals with $\alpha = 0.5$ are fair minded, as they put equal weight on payoffs to self and other.

Table 14: Average Treatment Effects on Business Outcomes

	(1) Log number of clients	(2) Log revenues	(3) Log expenditures	(4) Log profits	(5) Positive profit (=1)	(6) NHIF empaneled (=1)
AHME treatment						
Coefficient	0.260	0.339	0.036	1.573	0.082	0.145
Standard Error	(0.162)	(0.195)	(0.185)	(0.933)	(0.063)	(0.066)
p-value	[0.056]	[0.042]	[0.422]	[0.047]	[0.097]	[0.015]
Mean of control	387	3,340	3,225	167	0.71	0.21
Observations	187	184	182	180	187	187

Note: Standard errors in parentheses; one-sided p-values are in brackets ($H_0 : \beta \leq 0$); adjusted p-values for multiple hypothesis testing (Romano & Wolf) in angled brackets. Profits were transformed using the hyperbolic arc sine transformation $IHST(x) = ln[x + \sqrt{(}x2 + 1)]$ All business outcomes figures based on the last full operational month. Scale is the total number of clients, including all outpatient and inpatient (if applicable) visits. Total revenues was computed as the summation of income from out-of-pocket clients, insurance companies claims (including NHIF), and pharmacy sales (if applicable). Total expenses include payroll, health inputs (medicines, lab tests and reagents, consumables), utilities, rent, transportation, communication, cleaning, outsourced services (such as waste disposal), taxes, and other fees in the last full operational month. Profit is the difference between total revenues and total expenses. NHIF stands for the Kenyan National Hospital Insurance Fund.

Table 15: Average Treatment Effects on Efficiency Outcomes

	(1) Unit cost (ln)	(2) Unit cost of clinical/health staff (ln)	(3) Unit cost of non-health staff (ln)	(4) Unit cost of non-labor inputs (ln)
AHME treatment				
Coefficient	-0.224	0.015	0.090	-0.263
Standard Error	(0.127)	(0.120)	(0.064)	(0.153)
p-value	[0.040]	[0.551]	[0.919]	[0.044]
Mean of control group	10.85	2.71	0.30	7.82
Observations	182	179	179	182

Note: Standard errors in parentheses; one-sided p-values are in brackets ($H_0 : \beta \leq 0$). The clinical/health-staff category includes medical and clinical officers, nurses, matrons and midwives, lab technicians, pharmacists, nutriologists, and community health workers. Non-health staff includes administrative staff (accountant, receptionist, records officer) and support staff (guards, cooks, chefs, and drivers). Non-labor expenses include health inputs, utilities, rent, transportation, communication, and other non-labor fees. Unit cost was calculated by dividing the total expense in a given category by number of clients.

Appendix A. Supplement for AHME Program Details

The African Health Markets for Equity (AHME) program was designed and implemented through a consortium of four implementing partners: Marie Stopes International (MSI), Population Services International (PSI), PharmAccess Foundation, and the International Finance Corporation (IFC). AMHE was funded by the Bill and Melinda Gates Foundation (BMGF) and the United Kingdom's Department for International Development (DFID). Two organizations-Marie Stopes Kenya (MSK) through its "Amua" network and Population Services Kenya (PSK) through its "Tunza" network - served as franchisers for a set of health care clinics.

Social Franchising: Specific Service Components

The specific services offered as part of the AHME package's franchising included the following:

1. *Expanded Scope of Clinical Services:* Clinics that participated in AHME were supported to provide the following clinical services: family planning services, integrated management of childhood illnesses (IMCI), antenatal care and postnatal care (ANC and PNC), childbirth and delivery, diagnosis and treatment of sexually transmitted infections (STIs), and cervical cancer screening and preventative therapy (CCSPT).

2. *Improved Quality through Training for Clinical Staff:* AHME provided clinical training on national guidelines and standard operating procedures for the franchised health services.

3. *Increased Clinical Quality:* To help clinics adhere to best clinical practice guidelines, AHME implementing partners provided technical assistance to better organize service delivery and conducted regular clinic audits to supervise clinical procedures.

4. *Increased Demand for Clinic Services:* AHME supported clinics in developing community outreach plans using community health volunteers (CHVs). The CHVs educated community members about health services and referred them to franchised providers. CHVs were paid for effective client referrals. AHME clinics also advertised through call centers, social media, radio campaigns, and road shows. Franchised clinics were branded with franchise colors and logos that could only be used by franchise network members in order to identify clinics as ones that provided high quality services.

5. *Subsidized Commodities and Equipment:* Some contraceptive commodities were provided to the franchisees at free or subsidized prices. There was also a starter kit containing commodities to treat childhood illnesses.

6. *Monitoring and Reporting:* MSK and PSK franchisers oversaw the regular capture of data at clinics and performed data verification checks for data accuracy before data was sent to government agencies.

SafeCare: Specific Service Components

SafeCare was developed by an independent non-governmental organization (NGO) with three partners: the PharmAccess Foundation, the Joint Commission International (JCI), and the Council for Health Service Accreditation of Southern Africa (COHSASA). The AHME package's SafeCare program focused on the following 13 service areas:

1. *Governance and Management:* This area focused on strategic and financial planning practices. These practices included: whether clinics were properly licensed, whether strategic and operational plans existed and were being followed, whether a mission statement and plans for care and services that addressed national rules and regulations were available, and whether the organizational structure was documented. Record keeping was also assessed, along with financial and service auditing.

2. *Human Resource Management:* This area focused on the documentation of staffing plans, roles, and responsibilities, as well as the procedures for evaluating and verifying credentials of medical staff. Staff orientation and ongoing trainings were also reviewed.

3. *Patient and Family Rights and Access to Care:* This area assessed evidence of educating patients about their rights, maintaining patient privacy, and informing patients about their care, from initial consultation through treatment.

4. *Management of Information Services:* This area covered the use of a Health Information Management System (HIMS), as well as data quality for records if such a system was used. The privacy of the data kept, as well as whether the data were analyzed to improve the targeting of care to clientele, were also measured.

5. *Risk Management:* This area focused on the safety of patients and staff. Occupational health and safety, including sufficient personal protective equipment (PPE), infection prevention, and proper disposal of health care waste, were all covered.

6. *Primary Health Care Services:* This area focused on whether the clinic had adequate staff and infrastructure to deliver quality outpatient clinical services. This included medical equipment and supplies, as well as sanitation supplies and sterilization equipment. Documentation of and adherence to standard operating procedures (SOPs) across a wide range of common services were checked.

7. *Inpatient Care:* As with outpatient care, there was a focus on privacy, sanitation, and waste management. Documentation of and adherence to SOPs for routine and emergency procedures were also checked. In addition, continuity of care around the clock was assessed through indicators verifying that procedures and hand-offs were carefully documented and standardized across staff.

8. *Surgery and Anesthesia Services:* This service area measured the quality of surgical services for clinics that offered such services. This service area focused on staff qualifications and whether SOPs for anesthesia mixing and monitoring existed and were followed. This service area also included criteria for using pre-op and surgical checklists. Supplies for PPE and routine and emergency care stocking levels were recorded, as well as supplies and practices for sanitation and sterilization.

9. *Laboratory Services:* Quality of care and safety were the main focus covering staff, equipment, supplies, and protocols. This included whether staff were sufficient and qualified, whether there was enough PPE for the staff, and whether there was adequate laboratory equipment, supplies, and tests. Documented SOPs for lab tests, as well as internal quality control procedures for test kits, were checked. Waste disposal procedures were also assessed.

10. *Diagnostic Imaging Services:* Clinics that offered these services were assessed on available and sufficient equipment, ultrasound machines, and x-ray machines, as well as qualified staff to operate them. Documented SOPs and internal quality control procedures were checked.

11. *Medication Management:* This service area covered in-clinic pharmacies, as well as medications dispensed at clinics that did not have their own pharmacies. The criteria included qualified staff, adequate medication stocks, and infrastructure conducive to storing medicines, as well as systems and procedures in place to regulate drug procurement and distribution.

12. *Clinic Management Services:* This service area focused on building infrastructure, electricity, clean water, equipment maintenance, sanitation facilities, and information and communication technologies (ICT).

13. *Support Services:* This area covered food, laundry, cleaning, and waste disposal. The criteria focused on sanitation, infection prevention, and limiting contamination.

Each of the 13 SafeCare service areas contained between 6 and 37 criteria. SafeCare Coordinators assessed noncompliance, partial compliance, or full compliance for each of the indicators applicable at a clinic. Based on this assessment, clinics were assigned a SafeCare level between 1 (lowest) and 5 (highest). Each assessed clinic then worked with a SafeCare Coordinator to create a QIP that outlined areas to improve. The QIP targeted assessment indicators that were not fully compliant. According to the SafeCare consortium, Coordinators focus first on the issues that represent the highest risk to patients, visitors, and staff. SafeCare Coordinators on the ground have suggested that clinic owners help select the items in their QIPs based on the feasibility of implementation, with the idea that some quick wins through "low-hanging fruit" motivates clinic engagement. SafeCare Coordinators performed routine visits to the clinics, at least quarterly, in order to provide guidance and check on the progress of implementation of the QIP. Reassessment typically occurred every two years, though this varied across clinics, with SafeCare coordinators recommending reassessment based on perceived readiness. SafeCare routine visits continued for the duration of the AHME program.

Business Support: Category Components

The AHME package's Business Support focused on the following categories:

1. *General Business Operations:* This category assessed the clinic's governance structure, current legal registration status, whether clinics pay taxes, whether there are staffing and equipment plans, and whether a risk management plan is in place.

2. *Financial Management:* This category examined whether the clinic kept complete and accurate income and expense records, maintained functioning debt management processes, and utilized formal financial institutions as opposed to more informal systems and processes. Whether clinics had business bank accounts and the extent clinics utilized their business bank accounts and kept their personal finances separate from clinic finances was also tracked. This category also examined the extent clinics collected and utilized data on revenues and costs and whether business accounts were audited annually.

3. *Banking and Banking Records:* This category measured the extent to which clinics utilized formal financial institutions, opposed to more informal systems and processes. Whether clinics had business bank accounts were assessed, and the degree to which clinics utilized their business bank accounts and kept their personal finances separate from clinic finances were also assessed.

4. *Stock Management:* This category assessed whether drugs and supplies were systematically managed, costed, and audited. The two most important criteria were the extent to which clinics had a stock management system in place and whether expiration dates were recorded and monitored for all the drugs and supplies.

5. *Marketing and Demand Creation:* This category examined the thoroughness of clinic marketing plans. This included the degree to which these plans addressed goals to increase client utilization and revenues and informed surrounding communities about services offered at clinics, and whether clinics had methods to measure return on investment for marketing activities.

Appendix B. Supplement for Methods

B1 Experimental Design

Treatment Arm Recruitment and Control Honing

In the treatment arm, 373 clinics were dropped from the sample to streamline recruitment efforts and improve the statistical efficiency of the evaluation sample. These clinics were ranked "lower" than the last visited clinic in the first round of recruitment. The remaining 478 clinics were visited and screened by either MSK or PSK, and of these, 123 clinics that initially expressed interest in franchise services were offered enrollment by either MSK or PSK. These 123 clinics are considered the treatment arm evaluation sample for ITT analysis. Of these 123 "ever-franchised" clinics, 31 were defranchised by or did not maintain their relationship with MSK or PSK after initial recruitment.

The treatment honing process was replicated in the control arm by a specially trained evaluation team using a set of standardized operating procedures and field instruments. The random ranking of the last visited clinic in each county was determined based on the last visited clinic for the treatment arm in the same county. The ranking of the last clinic was used to establish an equivalent cutoff for dropping clinics in the control arm ("Not Visited": N = 336). Through this process, 493 control clinics were identified for evaluation honing team visits to replicate franchise recruitment procedures used for treatment arm clinics. Of these 493 clinics, 391 clinics were available and willing to participate and were visited for screening. Of these 391 clinics, 109 were considered eligible and interested in franchising and were included in the control arm evaluation sample.

B2 Program Compliance

To understand the extent of program compliance, we tracked which clinics received the AHME's four clinic-level interventions-social franchising, SafeCare, business support, and NHIF empanelment support-in clinics assigned to the "treatment group" (i.e., participating in the AHME program) and clinics assigned to the "control group" (i.e., not participating in the AHME program). We expected treatment clinics to receive the full package of AHME interventions, and the control clinics to receive none of the interventions. The extent to which a clinic received the interventions in the treatment group is the degree to which it "complied" with the impact evaluation's research design protocol, and the extent to which a control clinic did not receive the intervention is the degree to which the clinic complied with the impact evaluation's research design protocol.

In order to ascertain program compliance, three rounds of monitoring were carried out at a total of 123 clinics that were initially recruited in the treatment group of the AHME evaluation sample: two rounds were done at intermediate stages, and one was a part of endline data collection activities after the program ended. For the first two rounds, discrepancies were identified and referred to franchisers to assist them in improving coverage of and compliance with interventions. After the final round, we calculated the exposure window for each clinic (i.e., the duration each clinic participated in the various interventions), and we ascertained the frequency Social Franchising Coordinators visited and followed-up with a subset of clinics in their respective franchising networks to improve their performance on the interventions. We further identified clinics that had been assigned to the control group but ended up participating in one or more interventions.

More details, including figures and tables, on program compliance for clinics can be found in Gertler et al. (2020).

B3 Analytic Samples and Balance

This section describes the selection of the clinic, household, and quality of care samples used for analysis, as well as balance at baseline which allows us to assume exchangeable AHME treatment and clinic arms for our empirical ITT approach on endline data. We provide evidence supporting the effectiveness of randomization at creating clinic treatment arms that are, on average, exchangeable (balanced with respect to observable baseline characteristics). We provide similar results for the recruitment and sampling of the households and quality of care analytic populations from the 232 clinics in the AHME impact evaluation's clinic sample.

The structural quality analytic sample consists of 199 of 232 AHME evaluation clinics that consented and completed the clinic survey at endline. This section describes the clinic sample in detail. As for the analytic samples for process quality and health care outcomes, the SP data were collected at endline from 211 clinics belonging to the full AHME evaluation clinic sample (N=232). The client exit interviews, provider survey, household survey, and clinic survey data were collected at endline from the same 199 clinics as the structural quality analyses. Among the 21 clinics not included in the SP analytic sample, 8 were excluded, and 13 were closed. We excluded 8 clinics because: a clinic was located in a conflict region (N=1); no consent was given in previous AHME evaluation surveys (N=3); or clinics were ineligible to receive SP cases due to services provided (N=4: 1 eye clinic, 1 fistula clinic, and 2 clinics designated to serve employees from specific factories).

Clinic Analytic Sample

Of the 232 clinics in the final evaluation sample, 12 had closed before the endline survey (February 2019), and 14 were excluded for various reasons: unwillingness to participate in research (N=11); clinic could not be tracked down (N=1); clinic is now a public hospital (N=1); clinic located in high-risk area (N=1). Of the remaining 206 clinics visited for inclusion in the endline survey, and after explaining the research objectives and survey content to clinic managers, 199 consented to complete the first round of the survey.

Baseline Balance of and Description of the Clinic Sample

We compared key baseline characteristics across treatment and control arms for the clinic sample, including manager characteristics, services offered, and financial and operational indicators. Table 1.A reports treatment arm specific means, difference in means, and associated t-test results for select baseline clinic characteristics. Overall, we find that on average, clinic owners and managers are comparable across treatment arms. A large proportion of clinic owners also operate as managers (treatment: 67%; control 62%). Managers are comparable in terms of their qualifications (87% of managers in treatment clinics; 89% of

Appendix Table 1.A Balance of Clinic-level Characteristics at Endline by AHME treatment

	Control (n=91)	Treatment (n=108)	
	Mean	Mean	Difference
Number of patients (last week)	94.49	84.07	10.42
Ownership (years)	9.94	10.45	-0.51
Provides ANC	0.66	0.68	-0.02
Provides labour and delivery	0.39	0.40	-0.01
Provides PNC	0.61	0.64	-0.04
Provides child immunization	0.45	0.30	0.15**
Provides well-baby check-ups	0.72	0.61	0.11*
Provides TB treatment (adults)	0.06	0.10	-0.04
Provides inpatient services	0.22	0.18	0.05
Share of Family Planning clients	0.17	0.16	0.01
Profit margin (average 6 months)	0.38	0.41	-0.03
Has NHIF	0.07	0.14	-0.07
Business paid for employees	0.15	0.21	-0.06
Facility sells medicines to the public	0.48	0.50	-0.02
F-test of joint orthogonality			1.68*
F-test, number of observations			199

Notes: The values displayed in the last column are the differences in the means across the groups. The value displayed for F-tests are the F-statistics. All missing values in balance variables were imputed with the mean. ***, **, and * indicate significance at the 1, 5, and 10 percent critical level.

managers in control clinics had a clinical degree); demographic characteristics (gender, age); and experience (tenure, positions at other clinics) (see table 1.B).

Table 1.A reports means and standard errors for the types of services provided at the clinics, and other key financial and operational characteristics by treatment arm. Of the 199 clinics in the analytic sample, we find that clinics are similar on the following characteristics: utilization levels, number of years the clinic has been operating under the same management (ownership); and the proportion of clinics that provided services such as antenatal care, labor and delivery, postnatal care, tuberculosis testing and treatment, and hospitalization (inpatient services). Clinics in the sample also had similar unit costs, profit margins, and comparable levels of contractual arrangements with the government (National Hospital Insurance Fund, NHIF) and private companies. The only two differences that were found are related to the scope of child health services offered between treatment and control groups: the proportion of clinics that offered child immunization services (30% in the treatment group vs. 46% in the control group, p-value<0.05), and the proportion of clinics that offered well-baby check-ups (62% in the treatment group vs. 73% in the control group, p-value<0.10).

Quality of Care Sample

For the quality of care analyses, we used five data sources from different samples: (1) endline clinic survey; (2) endline data from standardized patient (SP) visits to clinics in the evaluation sample; (3) endline patient exit interviews; (4) endline household surveys; and (5) provider survey data collected among providers (e.g., doctors, nurses, clinical officers) identified during the SP visits. For all data sources, we recruited clinics and households in the clinic catchment areas based on the descriptions above for the full clinic and household evaluation samples. In the sections that follow, we further describe further the quality of care samples and inclusion and exclusion criteria for clinics in the endline SP and provider surveys.

Description of Quality of Care Sample

Analyses using the endline clinic surveys, patient exit interviews, and the provider survey rely on the full clinic evaluation survey sample. The household data analyses related to quality of care rely on the full household sample at endline. Analyses using SP data relied on clinics from the full evaluation sample that were both eligible to receive clients for our study's SP case scenarios and had consented to previous AHME evaluation endline surveys.

We collected quality of care measures from the endline clinic survey and patient exit interviews from the clinic sample (N = 199). The endline household survey relied on the endline household survey sample. As for the SP and provider surveys, figure 6.1 depicts the quality of care analytic samples by AHME treatment assignment for the SP and provider surveys. For the SP sample, we restricted the 232 randomized clinics from the full evaluation sample by excluding clinics that did not consent to previous surveys (N=3), were located in a conflict region (N=1), and were ineligible for SP cases (N=4). In total, 8 of the 232 clinics were excluded or ineligible for SP surveys. Between February and May 2020, after data collection for endline exit interviews was completed, SPs attempted visits at the remaining 224 clinics (treatment clinics: N=120; control clinics: N=104). All of these visits were conducted by SPs trained to portray pre-scripted cases seeking outpatient walk-in visits for family planning, child curative, and adult curative services.

During the SP data collection, supervisors and SPs worked to identify providers seen at the clinics. We constructed a provider survey sampling frame from the list of identified providers, and the data collection team was instructed to interview the full list. The group of providers that were not successfully identified were classified as "unidentified" providers. Among the treatment group, 651 SP visits seen by 221 identified and 128 unidentified providers were successfully conducted at 114 treatment clinics. Among the control group, 544 SP visits seen by 179 identified and 101 unidentified providers were successfully conducted at 97 clinics. We analyzed all 1195 successful SP visits from 211 clinics.[17]

For the endline provider survey, which was conducted in November and December 2019 and after SP data collection was completed, we excluded 8 treatment and 11 control clinics

[17]Of the 1195 successful visits, 1 was attempted 2 times before successful on the 3rd attempt; 20 were attempted 1 time before successful on the 2nd attempt; and 1174 were successfully conducted on the 1st attempt. The high rate of success on the 1st attempt was due to favorable notes taken during clinic mapping, but also largely due to the very low rates of provider absenteeism, much in contrast to what we've experienced conducting SP studies in some other settings.

Appendix Table 1.B Balance of Clinic Manager Characteristics at Endline by AHME Treatment

	Control Group (n=91)	Treatment Group (n=108)	Difference p-value
Manager is also the owner	0.62	0.67	0.454
Manager has clinical degree	0.89	0.87	0.644
Manager is female	0.34	0.36	0.765
Manager's age (years)	44.9	47.9	0.141
Manager's tenure (years)	9.17	9.44	0.801
Manager works at other private health facility	0.23	0.19	0.431
Manager works at a public health facility	0.11	0.08	0.528
Facility experience (years)	12.8	11.8	0.379

Note: Proportions reported, unless otherwise noted.

from the SP clinic sample (N = 211) based on exclusion criteria applied to the clinic survey sample. If providers were not identified or not available for an interview or could not be rescheduled for an interview, another provider who saw outpatient clients for family planning, child curative, or adult curative services was interviewed and served as an analytic replacement. In all, 322 providers were interviewed for the provider survey.

Baseline Balance of Quality of Care Sample

For our quality of care analyses, we relied on balance of clinic indicators collected at baseline for the clinic survey sample. We evaluated balance of the same clinic indicators collected at baseline for the standardized patient (SP) clinic sample. In this section, we report baseline balance of the SP clinic sample. These clinic indicators include client utilization, ownership, services provided, average costs and average revenue per patient, profit margins, whether the clinic accepted NHIF, whether the business paid for staff, and whether the clinic sold medicines to the public. For clinic indicators collected at baseline, we see that the N=211 clinics in the SP sample were balanced on 15 of the 17 variables at the 5% significance level. By chance, some characteristics were statistically different between the AHME treatment and control households, at the 5% significance level. Notably, whether the clinic provides child immunization (control proportion = 0.48, SE 0.05 vs. treatment proportion = 0.33, SE 0.04) and whether the clinic provides well-baby check-ups (control proportion = 0.74, SE 0.04 vs. treatment proportion = 0.61, SE 0.05) were significant at the 5% significance level. For this reason, we control for these two indicators in our analyses on household data.

Household Evaluation Sample

In this section, we outline the recruitment of households from the clinic evaluation sample and describe the creation of the endline analytic sample. We then utilize household characteristics collected at baseline to assess how the AHME household sample compared to a nationally representative sample of Kenyan households (restricted to regions where the AHME evaluation was conducted), and whether randomization created balanced treatment arms within the AHME analytic sample.

Household Sample Recruitment from the Clinic Evaluation Sample

In order to identify households associated with one of the 232 evaluation clinics at baseline, we conducted exit interviews at baseline. Of the 232 clinics in the AHME evaluation sample, 2127 households were recruited from 216 clinics at baseline. Eligible households could not be recruited from 17 clinics at baseline (treatment: N=8; control: N=9), because the clinics were closed or not operational during the exit interview period; the owner did not consent to clinic exit surveys; patient flow was limited and no eligible households were identified after 7 to 10 days of screening clients; or no households could be successfully located from information provided in the exit interview. At endline, 10 clinics were excluded from the clinic and quality of care analytic samples; 2095 baseline household interviews were associated with the remaining 206 clinics and defined the target panel household sample. At endline, 1295 endline household surveys were completed for 199 clinics, and for 7 clinics, none of the N=23 households could be tracked or contacted at endline. An additional 747 panel endline surveys were not completed (treatment: N=392; control: N=355) for the following reasons: randomly selected and excluded for logistical reasons (N=203); household moved and could not be located (N=170); household could not be located based on available information (N=125); household refused (N=56); household missed multiple appointments (N=88); other (N=110).

Description of AHME Household Sample

The AHME evaluation was conducted in seven Kenyan provinces: Central, Coast, Eastern, Nairobi, Nyanza Rift Valley and Western. We report how the households in the AHME analytic sample compared, on average, to a representative sample of households in these regions based on pre-intervention data collected from AHME households in 2015 and 2016. Demographic and Health Surveys (DHS) fielded the Kenya Malaria Indicator Survey (MIS) in 2015 NMCP, KNBS, and ICF (NMCP et al.). The 2015 DHS MIS survey (referred to as "DHS" from here on) was a nationally representative survey that collected data on household assets and demographics that overlapped with AHME household baseline survey indicators, and included the seven provinces where AHME was conducted (province was the smallest political unit that could be identified and linked between the AHME and DHS datasets). In addition to providing comparable household demographic and infrastructure characteristics that could be compared with the AHME sample, the DHS dataset also included a list of household assets could be used to create a nationally representative standardized asset wealth score and nationally representative wealth quintiles. Because AHME data were collected on the same set of assets, in the same time frame, the weights used to calculate the national wealth score could be applied to AHME households and asset wealth between the two samples could be directly compared. Specific details on the construction of the asset wealth index are available upon request.

Comparison of AHME and DHS Households in 2015/2016

Compared to DHS households, AHME households in the same province, that accessed care at AHME clinics in 2015 or 2016, on average, were slightly more likely to be urban (AHME: 42%; DHS: 41%) and to have more household members (median household members, AHME: 5; DHS: 3) (table 1.C). Heads of households were more likely to be male in the AHME sample (AHME: 89%; DHS 64%), younger (median years of age, AHME: 35; DHS 39), and more likely to have any education (AHME: 96%; DHS 87%). In 2015 and 2016, households in the AHME sample were also more likely to have a finished floor (AHME: 62%; DHS 58%) and to report having access to an improved source of water (AHME: 78%; DHS 70%) and their own, non-shared, improved sanitation facility (AHME: 47%; DHS 28%) (table DCL). The distribution of household asset wealth among the AHME and DHS populations based on asset ownership in 2015 and 2016 are similar; however, the AHME distribution is truncated at both tails, suggesting it has proportionally fewer "very poor" and "very rich" households compared to representative households in the same provinces. The increased peak of the AHME distribution also suggests a higher concentration of households that had median and above-average wealth at baseline, compared to DHS households in the same regions. We further found that the DHS households in the regions where AHME was rolled out followed the same quintile distribution as the national population (20% in each quintile). Households included in the AHME endline sample were less likely to be in the first (poorest) quintile (AHME: 13%; DHS 20%) and fifth (wealthiest) quintile (AHME: 10%; DHS 20%), and were more likely to have median levels of wealth, with an increasing trend toward the wealthier quintiles (quintile 2: 24%; quintile 3: 25%; quintile 4: 26%).

Taken together, at baseline households in the AHME household analytic sample appeared to have higher levels of wealth, education, and improved infrastructure compared to a representative sample of households from the regions where AHME was rolled out and who were sampled in the same time frame as AHME baseline. These findings suggest that clinics enrolled in the AHME sample attracted wealthier households at baseline, compared to the average household in that region.

Baseline Balance of Household Sample

Within the AHME sample, we evaluated the balance of demographic, socioeconomic, and infrastructure indicators collected at baseline. Table DCL reports the means (proportions) and standard errors for treatment and control households. Differences in means between the two groups are evaluated using t-tests for single indicators, and F-statistics test the joint statistical significance for each group of outcomes (e.g., demographics, infrastructure, socioeconomic indicators). We find that the endline household evaluation sample overall is well balanced on observable baseline characteristics, suggesting that randomization was effective. This provides evidence that the household control arm is a good counterfactual for the AHME treatment group and supports use of the ITT parameter throughout to provide unbiased estimates of the impact of treatment on household outcomes.

Appendix Table 1.C Balance of Baseline Household Characteristics in the AHME Household Analysis Sample

	(1)		(2)		(3)	(4)
	AHME Treatment		AHME Control		Difference (= T-C)	F-stat
Variable	Mean	(SE)	Mean	(SE)	Mean	
Demographics						
Urban setting	0.44	(0.05)	0.4	(0.06)	0.04	1.05
Number of HH members	5.07	(0.11)	5.1	(0.11)	-0.02	
Number HH members <5 years	1.29	(0.02)	1.29	(0.03)	0.00	
Male head of HH	0.89	(0.01)	0.9	(0.01)	-0.01	
Age of head of HH	37.69	(0.58)	37.06	(0.64)	0.63	
Head of HH's years of schooling	11.02	(0.27)	10.28	(0.32)	0.74*	
Expectant mother in HH	0.09	(0.01)	0.12	(0.01)	-0.02	
Infrastructure						
Improved sanitation	0.37	(0.03)	0.3	(0.03)	0.07*	1.23
Improved shared sanitation	0.32	(0.03)	0.34	(0.03)	-0.02	
Unimproved sanitation	0.31	(0.03)	0.36	(0.03)	-0.05	
Improved water source	0.81	(0.02)	0.75	(0.03)	0.06*	
Unimproved water source	0.19	(0.02)	0.25	(0.03)	-0.06*	
Natural floor	0.33	(0.03)	0.42	(0.04)	-0.08*	
Finished floor	0.66	(0.03)	0.58	(0.04)	0.08*	
Natural roof	0.03	(0.01)	0.03	(0.01)	0.00	
Finished roof	0.95	(0.01)	0.93	(0.01)	0.01	
Wealth Index[a]						
Wealth score (DHS)	0.3	(0.07)	0.18	(0.09)	0.13	1.28
Clinic Experience						
Attention received[b]	0.83	(0.02)	0.87	(0.02)	-0.03	0.91
Care received[b]	0.81	(0.02)	0.84	(0.02)	-0.03	
Wait time[b]	0.38	(0.03)	0.4	(0.03)	-0.02	
Clinic cost[b]	0.28	(0.05)	0.35	(0.08)	-0.07	
Transportation cost (in KSH)	52.73	(4.28)	62.81	(7.0)	-10.08	
Overall experience[b]	0.83	(0.02)	0.86	(0.02)	-0.02	
Insurance						
Any HH member has insurance	0.42	(0.03)	0.34	(0.03)	0.08*	1.8
Total HH members with insurance	1.44	(0.12)	1.25	(0.10)	0.19	
Share of HH members with insurance	0.32	(0.03)	0.26	(0.02)	0.06*	
Illness and Utilization[c]						
Child in household had fever	0.22	(0.02)	0.24	(0.02)	-0.03	1.5
Sought fever care	0.32	(0.03)	0.36	(0.03)	-0.04	
Child in HH: Upper respiratory illness	0.24	(0.02)	0.2	(0.02)	0.05	
Sought upper respiratory illness care	0.18	(0.01)	0.16	(0.02)	0.02	
Child in household had diarrhea	0.08	(0.01)	0.08	(0.01)	0.01	
Sought diarrhea care	0.04	(0.01)	0.03	(0.01)	0.01	
Households	715		580		Total N = 1295	
Clusters	107		92		Total clusters = 199	