

UC Berkeley

UC Berkeley Electronic Theses and Dissertations

Title

Managerial Practices and Altruism in Health Care Delivery

Permalink

<https://escholarship.org/uc/item/8458h7pj>

Author

Contreras Loya, David

Publication Date

2022

Peer reviewed|Thesis/dissertation

Managerial Practices and Altruism in Health Care Delivery

by

David Contreras Loya

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Health Policy

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Paul J Gertler, Chair

Professor Stefano M Bertozzi

Professor Aprajit Mahajan

Spring 2022

Managerial Practices and Altruism in Health Care Delivery

Copyright 2022

by

David Contreras Loya

(co-authored with Paul J. Gertler and Ada Kwan)

Abstract

Managerial Practices and Altruism in Health Care Delivery

by

David Contreras Loya

Doctor of Philosophy in Health Policy

University of California, Berkeley

Professor Paul J Gertler, Chair

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 a shift of provider preferences towards profit-maximization and away from altruism towards clients. The intervention also increased 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 7.8% (p -value < 0.001). 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 further examine intervention mechanisms using causal mediation analysis. We find that the effects on quality of care and price of services are mediated by the shift in provider preferences, while the increase in profits and reduction in non-labor unit costs are mediated by changes in management practices.

To Rita Ema Loya Chávez

To my mother, who continues to nurture and educate me by living an exemplary life example of grit and love. I am proud to continue your labor as a professor, and hopefully touch half the people you did over your school years. Thank you for teaching me how to dream bigger.

Para mi madre, quien continúa educándome y nutriéndome al vivir una vida llena de amor, pero también de fuerza y coraje. Estoy orgulloso de continuar tu labor como profesor, y espero llegar al menos a la mitad de las personas que tú tocaste durante tus años escolares. Gracias por enseñarme a soñar más grande.

Contents

List of Figures	iv
List of Tables	v
1 Introduction	1
2 The AHME Management Consulting Intervention	5
3 Methods	7
Experimental Design	7
Program Compliance	8
Data sources	9
Analytic Samples and Balance at Baseline	9
Empirical Strategy	10
Double Machine Learning	11
4 Results	12
Effect of AHME on Management Practices	12
Effect of AHME on Provider Altruistic Preferences	13
Effect of AHME on Quality of Care	16
Quality of Care Framework	16
Effect of AHME on Structural Quality	17
Effect of AHME on Process Quality	18
Effect of AHME on Perceptions of Quality of Care	21
Effect of AHME on Business Outcomes	21
Effect of AHME on Efficiency	22
5 Causal Mechanisms	23
Causal Mediation Analysis	24
Mediated proportions of altruism and management	25
6 Discussion	26
Ethical Review	28
Bibliography	46
1 Appendix 1. Supplement for AHME Program Details	51
Social Franchising: Specific Service Components	51
SafeCare: Specific Service Components	52

	Business Support: Category Components	54
2	Appendix 2. Supplement for Methods	54
	Experimental Design	54
	Program Compliance	55
	Analytic Samples and Balance	55
	Baseline Balance of and Description of the Clinic Sample	56
	Quality of Care Sample	57
	Description of Quality of Care Sample	57
	Baseline Balance of Quality of Care Sample	58
	Household Evaluation Sample	59
	Household Sample Recruitment from the Clinic Evaluation Sample	59
	Description of AHME Household Sample	60
	Comparison of AHME and DHS Households in 2015/2016	60
	Baseline Balance of Household Sample	61
	Construction of Economic Outcomes	61
3	Appendix 3. Supplement for Results (ITT difference in means)	65

List of Figures

1	Sample and Experimental Design	30
2	Evaluation Sample Locations	31
3	Treatment Exposure Time Distribution	32
4	Cumulative Distribution Function (CDF) of Overall Management (z -score) . . .	33
5	Cumulative Distribution Function (CDF) of profit-oriented preferences α	34
6	Cumulative Distribution Function (CDF) of Overall Structural Quality (z -score)	36
7	Cumulative Distribution Function (CDF) of Log of Price of Visit	38
8	Cumulative Distribution Function (CDF) of Number of Clients ($\ln$)	41
9	Cumulative Distribution Function (CDF) of Revenues ($\ln$)	42
10	Cumulative Distribution Function (CDF) of Profits (IHST)	42
11	Cumulative Distribution Function (CDF) of Unit cost ($\ln$)	44

List of Tables

1.1	Treatment Effects on Management Practices	33
1.2	Treatment Effects on Altruism	34
1.3	Quality of Care Conceptual Framework	35
1.4	Standardized Patient and Vignette Case Scenarios	35
1.5	Treatment Effects on Structural Quality	36
1.6	Treatment Effects on Knowledge of Correct Case Management	37
1.7	Treatment Effects on Investment	37
1.8	Treatment Effects on Correct Case Management and Price	38
1.9	Treatment Effects on Patient Satisfaction	39
1.10	Treatment Effects on Household Perceptions of the Clinic	40
1.11	Treatment Effects on Business Outcomes	41
1.12	Treatment Effects on Demand Generation Activities	43
1.13	Treatment Effects on Efficiency	44
1.14	Mediation Analysis of Altruism and Management	45
A1	Balance of Clinic-level Characteristics at Endline by AHME treatment	57
A2	Balance of Clinic Manager Characteristics at Endline by AHME Treatment	59
A3	Balance of Baseline Household Characteristics in the AHME Household Analysis Sample	62
A4	Revenue and Cost Categories with Descriptions and Examples	63
C1	Treatment Effects on Management Practices (ITT OLS)	65
C2	Treatment Effects on Structural Quality (ITT OLS)	65
C3	Treatment Effects on Knowledge of Correct Case Management (ITT OLS)	66
C4	Treatment Effects on Investment (ITT OLS)	66
C5	Treatment Effects on Patient Satisfaction (ITT OLS)	67
C6	Treatment Effects on Household Perceptions of the Clinic (ITT OLS)	68
C7	Treatment Effects on Business Outcomes (ITT OLS)	68
C8	Treatment Effects on Demand Generation Activities (ITT OLS)	69
C9	Treatment Effects on Efficiency (ITT OLS)	69

Acknowledgments

I want to thank my dissertation committee conformed by Paul J Gertler, Stefano M Bertozzi, Aprajit Mahajan, and Ziad Obermeyer for their tremendous patience, dedication, thoughtfulness, and words of encouragement. To Sergio Bautista-Arredondo, for onboarding me in AHME intervention during its early stages. To my fellow PhD colleagues Claire Boone and Nicole Perales for their conscientiousness, disposition, and tremendous amount of work in developing and analyzing the clinic survey. To Laura Chioda, for her friendship and for teaching me the meaning of excellence in conducting economic research. To Rita Cuckovich, whose invaluable and countless hours as a project manager made the difference in bringing multiple stakeholders together over the span of 6 years of program implementation. To Afke Jager, for her exceptional skills in research management and survey development, and for the emotional support provided during the launch of the endline clinic survey. To Jonathan Kolstad and Ben Handel, for their valuable feedback during the final stages of analysis and manuscript preparation.

To Innovations for Poverty Action Kenya, for their key role in data collection and support over the 6 years of program implementation (special thanks to Asman Suleiman, Allison Stone, Ginger Golub, Grace Makana, Eunice Kioko, Atanus Mutisya, and Vitalis Ogutu). To Nel Druce from the United Kingdom Foreign, Commonwealth and Development Office Department (FCDO, formerly known as DFID), Caitlin Mazzili at the Bill and Melinda Gates Foundation (BMGF), Kathryn Church and Anna Mackay from Marie Stopes International, Joyce Wanderi from Population Services Kenya, and Hans Peter Wiebling, John Dekker, and Annedien Plantenga at PharmAccess Foundation.

I am extremely grateful to BMGF and FCDO (formerly DFID) for funding this research. To Will Dow and Hector Rodriguez for their unconditional support. To Ada Kwan, my sister and profoundly admired colleague for always being by my side throughout this journey. Finally, to the Mexican National Council for Science and Education (CONACYT) for funding my doctoral studies at UC Berkeley.

Management and Altruism in Health Care Delivery

1 Introduction

Private for-profit providers are important players in health care markets in many low- and middle-income countries (LMICs)[32]. Moreover, the private sector is expected to grow rapidly as LMICs pursue universal health insurance that pays for care by private providers and significant funding gaps for publicly delivered care persist [72, 68]. However, improving and standardizing performance is a high priority as private sector quality of care is low and variable in many LIMCs [21, 19, 69, 62, 22, 20, 56, 23].

Private markets are typically comprised of many small clinics that are managed by clinicians with little or no management training or experience [7, 30]. Few employ modern management practices such as monitoring performance, financial audits, development of marketing and business plans, systems to analyze costs or profits, or use an exclusive bank account to manage financial transactions. The lack of sound management is concerning as it is critical to a clinics’s ability to efficiently translate inputs and clinical expertise into quality care, i.e., timely correct diagnosis of medical conditions and providing appropriate treatment.

We use a field experiment to evaluate the effect of a comprehensive management consulting intervention for private health care clinics in Kenya. The intervention was implemented by a consortium of NGOs with extensive experience in quality improvement and standardization as well as overall clinic business management. The intervention focused on management domains identified as being key: (i) standardizing operations, (ii) performance monitoring, (iii) goal setting, (iv) employee incentives, and (v) financial management [12, 71]. The consulting group first spent time at each clinic systematically accessing their business management and quality of care issues. Then, based on those assessments, developed specific improvement plans tailored to the needs of each individual clinic. The group then assisted each clinic to implement those plans. All clinics also received training in modern business and quality of care management practices.

Kenya is an important location for this study. The private sector owns and operates

42% of the 12,000+ primary care facilities.¹ Virtually all (98%) were managed by a medical professional with little business training or background; only 26% of clinics used a bank account that was exclusive to the clinic to manage financial transactions; and only 25% had ever taken a business loan. Private clinics treat an average of 67 clients per week, implying substantial excess capacity with clinicians waiting around for patients. There is substantial room for improvement in quality as only 52% of clinical cases were correctly managed and essential actions for correct case management varied widely [20]. Moreover, there could be gains from introducing simple professional management practices that are still lacking among providers such as financial audits, development of marketing and business plans, or systems to analyze costs or profits [7]. Finally the government is ramping up health insurance coverage [61, 47].

Recent evidence from field experiments supports the notion that management consulting can increase productivity in non-healthcare industries [14, 46, 15, 13]. It is less clear, however, whether this success easily translates to health care for a number of reasons. First, health care is highly heterogeneous combination of services that are tailored to each patient. Production is highly decentralized where patient care decisions are made by individual clinicians and not by clinics. Clinics influence patient care decisions indirectly through the structural environment of available diagnostic and treatment services, information and incentives.

Second, health markets are characterized by asymmetric information between providers and consumers where patients have trouble assessing quality [5]. Demand, then, tends to be quality inelastic and can lead to the under-provision of quality [29]. The management intervention, therefore, may not increase quality because profit oriented providers may not be willing to spend time and money to improve quality if it is not rewarded in the market with higher revenues as would be the case when demand is highly quality elastic.

Finally, a distinctive feature of health care is that many clinicians are, at least partially, motivated by altruistic preferences for patients and want them to receive high quality care [63, 10, 31, 37, 54]. This is important because, as we mentioned earlier, most private clinics are owned and managed by clinicians with little business training and experience. As such, the management intervention may be the first time that they discuss how to feasibly make large amounts of money from their clinic business. This could change owners view of their clinic from a place that delivers care to deserving patients to a place where they earn their living and make money. As such it might shift provider preferences to be more profit oriented. This is consistent with recent literature discussing how preferences are altered by experiences, opportunities, and events [1, 17, 11, 58]. If providers become more profit oriented, then the management intervention may actually lead to a reduction in quality especially in the presence of quality inelastic demand.

We show that the intervention statistically significantly improved management practices across all of the key domains. We find that after at least 2 years of implementation, the intervention was able to increase the adoption of better management practices. We measure management through a calibrated instrument combining audit-type and self-reported survey

¹Source: Ministry of Health, Kenya Master Health Facility List at <http://kmhfl.health.go.ke/>.

items. Our rich set of indicators map to the dimensions of management that have been documented to matter in health care: standardizing operations, performance monitoring, goal setting, and incentivizing employees [12].

We also find the the intervention increased provider preferences for profits at the expense of provider altruism. To elicit preferences, we use a real-stakes modified lab in the field dictator game. We adapt a dictator game originally based [4] and modified in several subsequent studies [57, 35, 6, 45]. In our case, the provider is given 19 different scenarios in which she decides how much of an endowment she wants to give to the client and how much she wants to keep for herself. We then use the data to estimate a CES utility function for each provider and recover an a provider-specific altruism parameter.

We then assess the impact of the intervention on quality of care using the standard structure-process-outcome framework [24, 28, 26, 25, 27]. While we find substantial improvements in the structural environment to be able to deliver high quality of care, we also find that process quality (appropriate diagnostic and treatment services provided) falls. We measure process quality of care using standardized patient (SP) audits [55]. SPs are healthy individuals recruited and trained to portray pre-designed health scenarios at sampled health facilities. Using SPs reduces the potential bias from patient heterogeneity since providers are visited by the “same patient” with the “same illness.”

Process quality is defined correct case management – the provision of essential diagnostic and treatment actions for specific conditions benchmarked against national guidelines. Overall, the intervention led to a *reduction* in correct case management of 7.8 percentage points (p -value < 0.05) relative to the control group mean of 63.6%. We also find that the intervention led to a 27% increase in the price charged clients for these services (p -value = 0.032).

Our results are consistent with the theory of imperfect markets in health care, as we show that 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. Moreover, we reject alternative explanations that these negative effects can be attributed to inadequate clinical knowledge or stock levels for necessary drugs supplies.

Despite our findings on quality of care, we find that clinics improved their business outcomes in several manners. First, 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. We attribute the increase in patients to the combined effect of outreach, participation of community health volunteers (CHVs) 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.

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 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 [5]. 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 an increase in adoption of management practices due to the AHME package of interventions. Second, the intervention shifted provider preferences in the direction of profit-maximization and away from altruism towards patients. Third, we document substantial improvements in structural quality, but no changes in provider knowledge. Fourth, we observe a reduction in process quality, specifically correct case management, which we define as the minimal and essential actions benchmarked against national guidelines. Fifth, client perceptions of quality and client satisfaction did not fall. Sixth, providers raised prices charged to clients. Lastly, AHME improved business profitability and efficiency.

The observed changes in clinical decision-making and clinic-level outcomes may mask heterogeneity driven by changes in social preferences among providers as a result of the management consulting intervention. 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: *How much of the average treatment effect is explained by changes in provider preferences and management?*

Using causal mediation analysis, we explored the mechanisms through which treatment affected outcomes. We tested the mediating role of provider preferences and management to explain the observed treatment effects, as these construct lie in the causal pathway between treatment and outcomes. We find strong and robust support of a mediating effect of preferences for profits on process quality and price of services. In contrast, experimentally-induced changes in management practices mediate the effects on clinic-level profits and unit costs.

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 [60, 59]. 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, our study contributes to an emerging literature in endogenous preferences formation where social interactions, opportunities, and experiences are able to alter a decision-maker’s utility weights when evaluating current and future outcomes. The influence of past experiences on behavior has a solid base in neuroscience, and so the economic and social implications of experience effects on decision-making are far-reaching, but matter specially in industries such as health care or education where intrinsic motivation is an important driver of effort.

Third, 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.

Lastly, 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 [38, 42, 36].

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, data sources, and our empirical approach. Section 4 presents results and detailed aspects of measurement, 5 provides a detailed explanation around causal mechanisms, and section 6 discusses our results.

2 The AHME Management Consulting Intervention

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 the implementing partners Marie Stopes Kenya (MSK)

under the Amua brand and Population Services 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.

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. 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 services² 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.

SafeCare: 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. To achieve this, the SafeCare 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 from which the partners developed customized Quality Improvement Plans (QIP) for each clinic. 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.³

Business Support: The main aim of Business Support in AHME was to enable owners to run their clinics as profitable professional businesses. The implementing partners first assessed clinic business practices using a business assessment tool (Appendix 1), from which they helped each clinic develop a Business Improvement Plan (BIP). 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 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), each containing

²See Appendix 1 for further details of franchising services.

³Further details on the service areas are detailed in Appendix 1.

3 to 7 criteria that were used to determine priority areas to improve business operations (Appendix 1). A Business BIP was drawn up 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. The partners also gave guidance on business development and facilitated financing for expansion by helping obtaining bank loans.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 121 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 232 primary care clinics that comprise the AHME evaluation clinic sample. We begin by describing the experimental design of the AHME impact evaluation, followed by a description of program compliance, and an overview of samples and types of data collected, including details on balance achieved at baseline.⁵ We conclude the section by discussing estimation methods, and details on measurement and variables construction are provided along with the presentation of results.

Experimental Design

The AHME field experiment was conducted across Kenya between 2013 and 2019 and consists of two main waves of data collection (baseline and endline) and tracking data across that time period. The experimental design is depicted in Figure 1 and below we summarize the process of sample construction and randomized assignment to treatment.

Before the AHME program began in 2012, we mapped all clinics in the evaluation study areas (35 of Kenya's 47 counties, see Figure 2) to define the sampling frame of clinics eligible for the program. 7683 clinics were first identified for mapping using various identified

⁴The actual loan take-up in our sample was less than 5%).

⁵Additional details on the larger impact evaluation are included in Appendix 2.

sources.⁶ From this universe of clinics, 4209 were excluded for various reasons, including being a faith-based or public clinic, being already franchised, or being near to an existing franchised clinic. The remaining 3474 clinics were screened by the research team using a standardized pre-screening survey that collected high-level clinic characteristics. Of the remaining clinics, 1800 were excluded for not meeting basic structural quality criteria such as lack of formal licenses or inadequate staffing. The remaining 1674 clinics were considered potentially eligible and were randomly assigned into two groups, treatment (n=849) and control (n=825). While clinics in these two groups met basic eligibility criteria, they still varied in their “level of eligibility” based on their existing capacity and suitability for receiving the AHME interventions. Both lists were then randomly ordered within strata defined by eligibility tiers within each county in the study.⁷

After randomization procedures were completed, we provided the implementing partners with partner-specific recruitment lists indicating the order in which clinics on their lists were to be approached for screening and recruitment.⁸ The implementing partners began engaging clinics on their lists and proceeded with their respective recruitment procedures in October 2013.⁹ Our power calculations showed that we needed to recruit 100 clinics on each arm, and the research team worked with the implementing partners to screen and recruit enough eligible clinics in the treatment arm. Eligible clinics that were invited to join the AHME program were considered part of the evaluation sample in the treatment arm. After completion of recruitment in the treatment group, the research team mimicked the same procedure in the list of n=825 potentially eligible clinics for the control group.

The final AHME evaluation samples were identified over the course of recruitment activities. For logistical reasons we recruited an additional 32 clinics over the original 200. In total, 232 clinics were identified for the final evaluation sample (treatment clinics: N = 123; control clinics: N = 109; see Figure 1). Take-up rates in treatment and control groups remained similar at 25 and 20%, respectively. In September 2016, baseline data collection, including baseline household and client exit interviews, was completed for all AHME clinics.

Program Compliance

Over the program implementation period, we tracked which clinics received the AHME’s four clinic-level interventions in clinics assigned to the treatment group and clinics assigned to the control group. The results of this monitoring shows that clinic-level implementation

⁶These 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.

⁷Three eligibility tiers (strata) were identified using data collected through the baseline survey instrument.

⁸During this second round of screening, we assessed whether the clinic kept business and health records, proximity to another franchised clinic, and their willingness to participate in the AHME program.

⁹Over this period, a number of clinics were unexpectedly closed because the government increased their licensing criteria. Such clinics were never visited by screening/recruiting teams.

fidelity of the AHME package of interventions was adequate. 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 member of the implementing partners' local organizations, and 74% were visited monthly or more often. While 80% had SafeCare and Business Support for more than 18 months, at least 90% of clinics in the treatment group had exposure to social franchising for more than 18 months. Approximately 76% were offered technical assistance with NHIF empanelment. Figure 3 shows the distribution of time of exposure to the program for the treated and control groups, which shows that $n=94$ of treatment clinics were exposed to AHME for at least 24 months, and that $n=19$ clinics originally assigned to the control group eventually received at least one of the three AHME intervention components for a period of 12 months or less. More specifically, $n=8$ control clinics were at some point franchised by either Tunza or Amua, $n=16$ received a SafeCare initial assessment and improvement plan, and $n=5$ received a Business Support initial assessment and improvement plan.

Data sources

The impact evaluation that accompanied the AHME program was designed to assess treatment effects on a number dimensions of health care performance: clinic operations, quality of care, and demand-side behaviors. To capture aspects that define health care structures, processes, and outcomes from different perspectives, we implemented five surveys at endline between 2018 and 2019: (1) a clinic survey; (2) an SP survey; (3) a provider survey; (4) client exit interviews; and (5) household surveys. Together, these surveys provide a comprehensive assessment of outcomes linked to the theory of change for AHME, which included management practices and clinic operations; financial sustainability; quality of care; utilization of clinics; client access to and utilization of health insurance for care visits; market impacts; impacts on vulnerable populations.

Analytic Samples and Balance at Baseline

Of the 232 clinics in the final evaluation sample, 11 had closed before the endline survey (February 2019), and 15 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 ($n=108$ AHME, $n=91$ control).

To measure health outcomes and market-wide effects of the program, we conducted baseline and endline household surveys. Exit interviews were used to identify households that were associated with the clinics in the evaluation sample. Of the 232 clinics at baseline, we

were able to recruit 2127 households from 216 clinics.¹⁰ At endline, 1295 households were successfully surveyed pertaining to the 199 of the 206 clinics that consented to participate in the survey and were operational at the moment of data collection.¹¹

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. Appendix 2 provides full details of the survey analytic samples and tests for balance at baseline for clinic, quality of care, and household analytic samples.

Empirical Strategy

We estimate treatment effects on different sets of outcomes. The first are related to management practices and provider preferences; a second set focuses on structural quality and provider knowledge; a third set on process quality, perceptions of quality, and the information asymmetries between providers and clients; followed by business profitability and efficiency. Lastly, we explore the mediating role of providers' social preferences and management in explaining the observed changes in outcomes as a result of the AHME program. We estimate treatment effects using partially linear models, where controls were optimally selected using Double/Debiased Machine Learning methods [18]. When the unit of observation was the clinic (e.g. management, profitability) we estimated the following model:

$$Y_i = \beta_0 + \beta_1 D_i + g(X) + \varepsilon_i \quad (1.1)$$

where Y_i is the outcome variable for clinic i , D_i is a dummy variable indicating whether or not the clinic was assigned to the AHME group, β is the treatment effect, $g(X)$ is an unknown but estimable function using Machine Learning, X denotes a set of baseline covariates, and ε_i is the error term.¹²

When the unit of observation was the client-provider interaction, the general model we estimated was:

¹⁰Eligible households could not be recruited from 16 clinics at baseline (treatment: $n = 7$, control: $n = 9$) because 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 interviews.

¹¹No households could be tracked or contacted at endline for $n=7$ clinics ($n=23$ households). The remaining 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 ($n=110$).

¹²Where necessary, we conducted additional sensitivity analyses to account for the effect of measurement error in self-reported financial data, such as running models with winsorized version of the dependent variable.

$$Y_{ij} = \beta_0 + \beta_1 D_{ij} + g(X_j, Z_{ij}) + \eta_i + \varepsilon_{ij} \quad (1.2)$$

where Y_{ij} is the outcome variable for individual i and facility j , D_{ij} is a dummy variable indicating whether provider i in clinic j belongs to the AHME group, β is the treatment effect, $g(X, Z)$ is estimated with Machine Learning, X_j is a set of facility-level controls, Z_{ij} is a relevant set of provider-level control and case variables, η_i are SP fixed effects, and ε_{ij} is the error term. The standard errors were clustered at the provider level to account for any intra-cluster correlation. All estimations were done in Stata using the command *xporegress*.

Indicators measured in monetary units were transformed using the natural logarithm and the inverse hyperbolic sine [16]. 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 followed [9]. 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 [67]. The clustering pattern was factored in for the cross-fitting process, bootstrap process, and inference procedures where applicable (e.g. SP or household-level data).

Double Machine Learning

We use Double/Debiased Machine Learning (DML) to estimate causal parameters for two reasons: to increase precision of ITT estimates, and to investigate the causal mechanisms through which AHME affected outcomes.

DML is a method for estimating heterogeneous treatment effects in designs where the assumption of selection on observables holds, but the design matrix is either high-dimensional or the functional form of the nuisance function is unknown and cannot be adequately modeled using standard parametric approaches. DML relies on Neyman-orthogonality conditions to impose a double robustness property with respect to model mis-specification of outcome and treatment. The use of sample splitting (cross-fitting) is done to prevent overfitting. To achieve root- n consistency and conducting valid inference, we will follow [18] whose work shows that the estimation strategy satisfies a specific moment condition (linearity) and Neyman orthogonality of the score. This implies that the score function is smooth in the nuisance parameter and the estimand (η, Ψ_{d0}) and therefore, the plug-in estimators converge at a rate of $n^{-1/4}$ to achieve $n^{-1/2}$ convergence of the target parameters to their true value. It is possible to use any high-quality ML algorithm to achieve $n^{-1/4}$ convergence rates, and the

Stata implementation of DML we used relies on cross-validated lasso as the regularization method to learn the function $g(X)$.

Our models rely on a rich set of controls at the provider, facility, and county level. Provider-level controls include gender, qualifications, a knowledge index derived from medical vignettes, and tenure time at the clinic. Facility controls include clinic age, whether the manager is the owner, a baseline management score, indicators for the presence of laboratory and inpatient facilities, share of family planning clients, average margin profit over a six-month period, and the average weekly client load. At the county level, we include indices of socioeconomic status, burden of disease, and availability of health services. Results of treatment effects estimated as difference in means of the ITT parameter are shown in Appendix 3.¹³

4 Results

Central to AHME’s supply-side strategy was a package of interventions designed to improve managerial practices and scale-up private for-profit clinics across Kenya. Drawing from the literature on health markets failures and recent evidence on experience effects, we hypothesize that the program also led to behavioral changes in clinical decision-making by making the link between quality, profitability, and demand for care more salient in the clinic’s objective function. In other words, teaching clinicians how to make money altered their social preferences, and we use this endogenous source of variation to explain the observed changes in process outcomes.

This section reports treatment effects of AHME on the adoption of management practices, clinicians’ social preferences, structural quality (including provider knowledge and investment), followed by the results on process quality of care and perceptions of quality and satisfaction with care. We then present the clinic-level effects on business profitability, demand generation activities, and efficiency. To explain causal mechanisms, we report the role of management and provider preferences as mediators in the relationship of AHME with decision-making and clinic-level outcomes. Each subsection below describes aspects of measurement before presenting quantitative results.

Effect of AHME on Management Practices

To measure management practices, we grouped practices into eight general dimensions of management spanning clinical and business responsibilities: general governance, financial management, marketing activities, human resource management, supply chain management, clinical care management, business and financial operations information technology, and clinical information technology. We defined these categories by drawing from Bloom and van Reenen’s methodology as well as from implementing partners’ accreditation and business improvement standards. Since low- and middle-income countries lack a validated tool to

¹³All continuous variables were winsorized at percentiles 1 and 99, and were transformed to z -scores.

measure management in primary care clinics, we relied on face validity (subjective assessment of whether a test measures the content it purports to measure) to group survey items. Although we find wide variation in the internal consistency of our dimensions, our overall management instrument is a reliable measurement tool.

To construct indicators management practices at endline, we used observations by the enumerators administering the survey and responses from clinics recorded in the clinic survey. We designed our instrument and derived several clinic survey items from implementers' internal 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), which 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. Our final indices map to the categories of management that matter in healthcare: Standardizing operations, performance monitoring, target setting, incentivizing employees, and financial operations [12]. These indices were computed as standardized weighted averages following [2], with respect to the mean and standard deviation of the control group.

The AHME intervention led to sizable and significant effects on the adoption and implementation in four out of five dimensions of management practices, as shown in Table 1.1. The largest effects were concentrated in financial operations (column 2), operations management (column 1), people management (column 5) and target setting (column 4). AHME clinics increased their compliance with financial operations in about 0.471σ relative to the control group (p -value = 0.002); 0.349σ for operations management (p -value = 0.009). For people management, the estimated effect is 0.295σ (p -value = 0.022), and 0.253σ for target setting (p -value = 0.069). The effects on performance monitoring (column 2) and on the overall management index (column 6) were positive but estimated without precision at conventional levels. When looking at the overall management score, the estimated effect is 0.386σ relative to the control group (p -value = 0.007). When plotting the cumulative distribution function (CDF) of the overall management index by treatment and control, we notice that the distribution of AHME is shifted to the right throughout its entire support (Figure 4), which is evidence of stochastic dominance and that the treatment effect is therefore not driven by the presence of outliers.

Effect of AHME on Provider Altruistic Preferences

Health professionals working in the private sector have a unique set of values that drive their care decisions in the form of social preferences, which are shaped by self-selection into health care careers and by their training.¹⁴ Health care professionals are also extrinsically motivated by money [63], but market conditions and lack of management training may prevent them from running profitable businesses. Therefore, when a program teaches clinicians and clinic owners basic management tools that enable them to better link inputs to out-

¹⁴Examples of values in the medical profession are the Hippocratic Oath, or the professional principle to do "no harm".

puts, profitability becomes more salient in the firm objective function. Our argument is that management training can alter managers' preferences by exposing them to a different set of realized experiences. Over time, clinic owners learn that it is possible to be more efficient by cutting down on costs that do not contribute to signaling more quality, or to increase revenue by rising prices or providing unnecessary care. More formally, clinicians hold mixed worldviews that they use to evaluate future choices. In other words, the clinician cares about her future utility, but also evaluates the experience producing that level of utility using her current worldview. A pure worldview in our setting lies between the two extremes of a continuum between altruism (saving lives no matter the cost/clients' ability to pay) and profit-maximization (exhausting clients' willingness to pay for quality). Therefore, our behavioral model endogenizes clinicians' preferences [11, 1]. In order for decision-makers to alter their preferences, one worldview must necessarily happiness-dominate the other, i.e. because the profit-oriented worldview yields higher utility irrespective of the choice taken (i.e. low quality, high quality).

Since theory says that clients are unable to recognize different levels of quality of care, we conjecture this to 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. Therefore, when two quality vectors have the same cost we assume that the clinic's decision maker chooses that quality vector which maximizes quantity of services demanded at a given price [63].

Following literature suggesting that provider preferences play a role in decision making among health care providers, we examine provider preferences, specifically distributive social preferences.¹⁵ 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 [35].

Measurement: To elicit altruism, we use a real-stakes modified dictator game. Dictator games provide a nonstrategic environment to elicit social preferences. We adapt a modified dictator game originally modified from [4] in several subsequent studies [57, 35, 6, 45]. In our case, the dictator is a provider at one of the clinics in the AHME evaluation sample and is given 5 different scenarios of real or fictitious clients with different characteristics (e.g. married, poor). The dictator then determines how much of an endowment she wants to give to the client and how much she wants to keep for herself.

The game combines decisions in zero-sum and positive-sum strategies, in the range of 100 to 500 Kenyan shillings (KSH, or about 1-5 USD). Depending on the scenario, we apply 1x to

¹⁵Distributive 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).

5x multipliers to the recipients' potential payoff. In this literature, the larger the multiplier, the more the dictator is likely to give. Over the course of the game, the dictator makes a total of 19 decisions, with each provider facing the same set of decisions which are spread across five different case scenarios: real client, fictitious childhood diarrhea client, fictitious family planning client, fictitious malaria client, and fictitious poor malaria client. Since each dictator divides the constant endowment of 1,000 KSH between themselves and the recipient, we can express their budget constraint as $\pi_s + p_0 \times \pi_0 = w$, where π_s is the payoff to self; π_0 is the payoff to others, and p_0 takes on values 1 to 5. When p_0 is larger than 1, the provider faces a trade-off between equity and efficiency.

After confirming our data is consistent¹⁶, we examine a preference parameter for altruism estimated following [4, 57]: $\alpha \in [0, 1]$, 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. We estimated the altruism parameter using a constant elasticity of substitution (CES) utility function, because it allows to identify the altruism parameter in a convenient manner. The estimates were generated using nonlinear Tobit maximum likelihood in Matlab.¹⁷

To identify treatment effects on provider preferences, we performed a grid-search approach by sequentially dichotomizing the providers according to thresholds of alpha: $\alpha \geq 0.5$ for fair-minded providers, and $\alpha \geq 0.75$, and $\alpha \geq 0.9$ for pure profit-maximizers.

We now turn to program effects on provider preferences. We find that on average, real clients received 48% of the budget share; all non-poor (real and fictitious) clients received 51% of the budget share; and poor (fictitious) malaria patients received 13% more budget share than non-poor (fictitious) malaria patients. With respect to the effect of the program on provider preferences α , our first result is that the CDF of α by treatment status (Figure 5) shows a shift in the distribution particularly between [0.4, 1], which corresponds to the range of providers who are fair-minded to profit-maximizers. Interestingly, we see that that providers at AHME control groups are more heavily concentrated around fair-mindedness

¹⁶To estimate the social parameters of interest, we performed checks on our data to confirm that each respondents' choices are consistent with individual utility maximization. If the budget sets in our modified dictator game are linear, we can invoke Afriat's Theorem, which states that if a finite dataset generated by an individual's choices satisfies the Generalized Axiom of Revealed Preference (GARP), then the data can be rationalized by a well-behaved (i.e., piecewise, linear, continuous, increasing, and concave) utility function [57]. We assessed the data's the degree of compliance with GARP by calculating Afriat's critical cost-efficiency index (CCEI) (Fisman et al. (2017); Li, Dow, and Kariv 2017). CCEI measures the amount the budget constraint should be adjusted to remove GARP violations. CCEI takes on values [0,1] and when CCEI is closer to 1, the smaller the adjustment of budget sets required to remove all violations, thus closer to satisfying GARP.

¹⁷The CES utility function is given by $u_s(\pi) = [\alpha(\pi_s)^\rho + (1 - \alpha)(\pi_0)^\rho]^\frac{1}{\rho}$.

($\alpha = 0.5$), and providers at AHME treatment groups are more heavily concentrated near perfect profit-maximization.

Table 1.2 shows the effects of the program on the likelihood of being located above three different thresholds of α . While there is no effect on belonging above the fair-mindedness threshold of 0.5 both in terms of the point estimate and its standard error (column 1), we observe that the program increased the share of providers with $\alpha \geq 0.75$ and $\alpha \geq 0.90$ by 12.7 and 9.4 percentage points, respectively, and the effects are statistically significant at the 5% level. These effects remain statistically significant with and without the inclusion of case, provider, and facility controls (results not shown).

Effect of AHME on Quality of Care

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 SPs.

Quality of Care Framework

To understand the impact of AHME on quality of care, we draw from Donabedian's conceptual framework for quality of care (see Table 1.3) and construct metrics based on national health guidelines [24, 28, 27]. 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); and the clinic (including market conditions 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 enter in the production function of process quality and are complements of clinical decision-making. 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 [28].

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.

Effect of AHME on Structural Quality

To capture structural quality, we followed a similar process as we did for management. First, we collected 3,178 indicators from monitoring data collected and reported by SafeCare, MSK, and PSK. We applied IRT to reduced this number to 32 internal monitoring indicators. 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 these constructs through clinic administrator responses and enumerator observations in the AHME endline clinic survey. Such indices were computed according to [2].

For investments in equipment, a dedicated survey module asked for the existence, quantity, and value of a curated 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 did not consider 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 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, 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 1.4 provides details of the case scenarios and the definition of correct case management.

We then turn to the effects on structural quality. We present DML estimates of two indices for structural quality compliance: interpersonal structural quality (16 of 33 indicators) and technical structural quality (17 of 33) and an overall index.¹⁸ 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. Table 1.5 presents treatment effects expressed as standard deviations with respect to the control group, where we find positive effects across these dimensions. Specifically, AHME increased compliance by 0.281 σ (p -value = 0.026) for interpersonal quality and 0.150 σ (p -value=0.183) for technical quality. The effect on the overall structural quality index was estimated at 0.521 σ (p -value < 0.001), and its CDF (Figure 6) shows stochastic dominance of AHME over the control group.

For provider knowledge, we find knowledge to be high for diarrhea and malaria with AHME control clinic means at 91% and 99%, respectively (mean of control group in columns 1 and 3 of table 1.6). However, knowledge was particularly low for correctly managing the family planning case (column 2): 11% of the providers showed correct knowledge 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 family planning SPs knew to do at least one of the four actions for 94% of the interactions (not shown). Overall, we find that the intervention did not change provider knowledge, as the point estimates treatment effects are statistically indistinguishable from zero including the overall knowledge index (Table 1.6, column 4).

To explore tangible changes by which the AHME program improved structural quality, we present the treatment effect on capital investments in table 1.7. The AHME program increased investment in medical equipment by 44% (p -value=0.063) and total investment by 42% (p -value=0.092). We also find a sizable but imprecisely measured reduction in investment in diagnostic/laboratory equipment of about 20%.

Effect of AHME on Process Quality

Measurement: Our study utilized the SP method, which can provide an unbiased and scientifically valid technique to answer questions related to provider practice. The SP method is increasingly considered the gold standard for measuring provider practice in a one-time en-

¹⁸In 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.

counter in countries where medical records are poor or absent [65, 53, 55]. 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 [55].

To measure process quality, we collected and analyzed primary data from $n=1195$ SP visits at endline. 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.¹⁹ Described in table 1.4, all scenarios except the asthma case were implemented as a vignette so that we would be able to compare knowledge versus practice. 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 2.

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 1.4. Correct case management definitions refer to minimal and essential actions as benchmarked against national guidelines for case management [49, 51, 48, 50, 52]. 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.²⁰

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

¹⁹We 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.

²⁰A 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.

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. 1.4 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.²¹ 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.

Table 1.8 reports the effect of the AHME program on correct case management from SP data for diarrhea (model 1), family planning (model 2), asthma (model 3), malaria (model 4), and all cases pooled (model 5). We observe that SPs seen at clinics receiving the AHME treatment had lower rates of correct case management: 7.8 percentage points (p -value < 0.001 ; model 5) on average, compared to the control group (mean = 0.636, relative reduction of 12%). When examining the different cases (models 1–4), the effect of AHME is large and negative for each case, but estimated without precision at $\alpha = 0.05$ level except for the diarrhea and malaria scenarios (15.2 and 10.5 percentage points, respectively).

Data on the price of health services were collected in several surveys, but our main source is the information captured by the SPs which is the effective price paid for the visit that generated that provider-SP interaction. The results on price of visit (table 1.8, model 6) shows the effect of AHME on the price of visit where we find that relative to an average of 340 KSH in the control group (about 3.4 USD), SPs seen by providers in AHME clinics were charged 26.9% more (p -value = 0.032). The CDF of log of price (Figure 7) shows stochastic dominance of AHME in the low and middle range of the distribution.

²¹Kenya 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.

Effect of AHME on Perceptions of Quality of Care

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 complemented our understanding by administering household surveys in clinic catchment areas to understand household perceptions of clinic quality.

Table 1.9 shows the effects of AHME treatment on client perceptions of clinic amenities from the perspective of SPs (upper panel) and real clients (lower panel). First, we find evidence supporting the hypothesis that AHME improved amenities, based on our 5-item index (coefficient = 0.102, p -value = 0.055; model 1).²² We do not find any significant program effects on whether SPs or clients found providers to spend sufficient amount of time at $\alpha = 0.10$ significance level. With respect to whether they completely trusted the provider’s medical treatment decision, we find no effect from the perspective of SPs but we do show a small but statistically significant effect from the perspective of real clients (coefficient = 0.029, p -value = 0.047).

Further, we find no effect of AHME on household perceptions of clinic quality. As shown in Table 1.10, clients of AHME clinics at baseline did not rate those clinics higher than comparison clinics (p -value = 0.620; column 1). Clients in the AHME treatment group were also not more likely to rank the AHME clinics as higher quality than clients in the control group (p -value = 0.326; column 2). Therefore, we observe that clients do not recognize the change in reduced correct care in the market and it demonstrates a market failure.

Effect of AHME on Business Outcomes

We collected clinic-level information about number of clients, revenues, costs, and profits for each clinic’s last full operational month. Endline clinic survey questionnaires included data extraction forms for capturing quantitative information of utilization (client load) and financial indicators. Data sources included Ministry of Health reporting forms, electronic databases, electronic records, electronic or written medical records, logbooks, registers, receipts, stock cards, clinic accounting records, payroll records, and performance reports.

Clinics’ financials were collected retrospectively. We reduced the cognitive load on respondents by using an iterative triangulation method with automatic consistency checks, based on previous work in Uganda [3]. This method aggregated different revenue sources, as well as detailed cost categories (see Appendix 2). 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 give their “best estimate” instead. Inputs the clinic received by donation were valued at their unit costs determined by local price quotes.

²²This 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.

Moving to several dimensions of business performance (Table 1.11), we find a substantial increase in the total number of clients who demanded health services in the last full operational month (column 1). Treatment clinics have, on average, 29% more clients (p -value = 0.042) relative to control clinics, which had 385 clients on average. The CDF of number of clients (Figure 8) shows the treatment effect to be concentrated in the low and middle range of the distribution, though we do not reject the hypothesis that no group stochastically dominates the other (Kolmogorov-Smirnov test p -value = 0.130).

The intervention led to an increase in total revenues (table 1.11, column 2). On average, treatment clinics had 38% higher revenues (p -value = 0.026) relative to control clinics, which reported US\$3,340 in the last full operational month. We also show that total profits increased in treatment clinics about 94% (p -value = 0.040) relative to an average of USD 167 in control clinics (column 3). Again, the CDF of total revenues (Figure 9) for the treatment group is shifted to the right in the low and middle range of the distribution, but we cannot reject stochastic dominance (Kolmogorov-Smirnov test p -value = 0.150). For profits, Figure 10 shows a difference in the empirical distribution between AHME and control in the low and middle range of the distribution of profits.

What types of mechanisms are able to explain the observed effects on business profitability? Table 1.12 shows the treatment effects on demand generation activities. We find that AHME led to a sizable increment in the proportion of enrolled clinics under the NHIF scheme (column 1). Compared to 21% of control clinics operating with NHIF, the proportion of AHME clinics reached 36.8% (p -value = 0.007). The share of clinics performing any type of marketing activities more than doubled as a result of AHME (column 2), relative to 35% in the control group (p -value < 0.001). With respect to the number of different marketing activities performed ²³ (column 3), we find an increase of about one activity, relative to 0.57 activities on average in the control group (p -value < 0.001). When looking at the share of clinics that have a community health volunteer worker as a demand generator (column 4), we find that relative to 20% of control clinics, the AHME program increased this share by 29.3 percentage points (p -value < 0.001).

Effect of AHME on Efficiency

Taken together, the estimated effects on number of clients, revenues, prices, and demand generation are consistent with our results on different measures of efficiency, as shown in Table 1.13. Column 1 shows that the program led to a reduction in unit cost of about 4.2 USD, relative to 10.4 USD in the control group (p -value = 0.050); this reduction is mainly driven by non-labor costs (column 3) as we show that AHME clinics spent 3.9 USD less than their control counterparts (p -value = 0.057). The increase in demand among AHME clinics led to an increase in the number of clients seen per clinical provider (column 4) of about 25% (p -value = 0.070). When looking at the number of different medicines that were

²³Respondents could select multiple options from the following list: distribution of promotional material, radio spots, sending SMS, vaccination campaigns, reproductive health day/week, HIV status awareness day/week, community health workers, medical services for free, or social media.

out of stock in the last year (column 5), we find that relative to an average of 1.2 items in the control group, the treatment effect of AHME is small and statistically insignificant (coefficient: -0.14, p -value = 0.347). AHME clinics are also purchasing laboratory inputs (column 6) at a price that is 17% lower than control clinics (p -value < 0.001), while for medicines (column 7) the estimated effect is statistically zero. The CDF of unit cost (Figure 11) shows that the distribution of AHME is shifted to the left in the middle and upper range of the distribution, though we cannot reject the null hypothesis of no stochastic dominance (p -value = 0.130).

5 Causal Mechanisms

After describing the reduced-form effects of AHME on the outcomes of interest, this section presents an exploration of intervention mechanisms. We begin by briefly introducing mediation analysis, the assumptions required to attribute causality to model parameters, and estimation procedures, followed by results showing the decomposition of the treatment effect into a direct effect of AHME and its indirect effects via changes in management and provider preferences.

The goal of mediation analysis is the identification and evaluation of the mechanisms through which a treatment affects outcomes. Empirically, the challenge is to disentangle the treatment effect into two components. The first component is the indirect effect, which is attributable to mediators, which are one or more variables that lie on the causal pathway between treatment and outcomes and are thus affected by treatment. The second component is the direct effect, which summarizes all other possible mechanisms for which a treatment affects outcomes.

Mediation analysis has a long history in epidemiology, psychology, sociology, and political science. Applied economists have not widely adopted this framework because of problems of identification of economic causal parameters unless strict independence conditions hold, with a major example being the widely used decomposition methods such as Blinder-Oaxaca that typically disregard endogeneity [40]. As pointed out by [8] and [66], the conventional exogeneity conditions are not even guaranteed in the randomized-controlled trial setting. More recently, [43, 44] unpacked the conditions under which traditional regression analysis can yield causal estimates, which essentially rule out confounding of treatment-outcome, treatment-mediator, and mediator-outcome²⁴.

Despite the evident appeal of mediation analysis in economics, no consensus has emerged on the best methods for estimating indirect and direct effects. Estimation methods can be traced back to structural equations modeling (SEM) and path analysis, but more recently the emergence of methods under the the potential outcomes framework include inverse probability of treatment weighting (IPTW) and double robust methods using machine learning [33, 41].

²⁴Imai et al. showed that structural equations rely on untestable assumptions in order to attribute a causal interpretation to model parameters, nor do they account for treatment-mediator interactions.

Causal Mediation Analysis

Let us introduce some notation. Our target parameter (ITT) described in the methods section can be expressed as $ITT = E[Y(1, M(1)) - Y(0, M(0))]$, or equivalently as the sum of the direct and indirect effects. The natural direct effect $\tau(d)$ represents the mean difference in potential outcomes as treatment (D) ranges from 0 to 1, keeping constant the potential value of the mediator M . The (average) indirect effect $\delta(d)$ equals the difference in potential outcomes as the value of M varies, while keeping assignment to treatment fixed.

The expressions for the average direct and indirect effects are thus:

$$\tau(d) = E[Y(1, M(d)) - Y(0, M(d))], \quad d \in \{0, 1\} \quad (1.3)$$

and

$$\delta(d) = E[Y(d, M(1)) - Y(d, M(0))], \quad d \in \{0, 1\} \quad (1.4)$$

The decomposition of the ITT (or the average total treatment effect, in the language of mediation analysis) as the sum of the natural direct and indirect effects allows for heterogeneous treatment effects across treatment groups, given the interaction between D and M .

Identification of $\tau(d)$ and $\delta(d)$ is based on the standard assumption of sequential ignorability [66] which posits that the treatment is as good as random given pre-treatment covariates, and that the mediator variable is also ignorable given the treatment assignment and pre-treatment covariates, denoted by X . This set of variables must not contain any post-treatment variables that confound the relationship between the mediator M and the outcome Y . In other words, treatment assignment D , the mediator variable M , and the outcome Y should be causally affected by distinct and orthogonal sets of unobservables. Therefore, none of the unobservables may jointly affect two or all three elements (D, M, Y) conditional on X [33]. Taken together, these assumptions rule out the presence of pre- and post-treatment confounders of the treatment-mediator and mediator-outcome relations.

In our study, conditional independence of the treatment holds by virtue of the experimental design. For conditional independence between mediator and potential outcomes, the concern is that the relationship may be confounded by unobserved post-treatment variables. The effect of AHME on quality and clinic-level outcomes depend on their levels of management and altruistic preferences, and potential confounders include providers' ability and knowledge, staff composition, scale of operation, and clinic profitability. These post-treatment confounders are causally related to different levels of the mediators and outcomes. To the extent that AHME affected such clinic and provider characteristics and are not controlled for in the model, this endogeneity issue will yield biased and inconsistent measures of the indirect effect.

We argue that our rich set of covariates allows to decompose and conduct valid inference on the effect of treatment via its mediators using data-adaptive estimation methods. In this sense, DML allows to flexibly and optimally control for observed baseline covariates, as they

proxy for potential treatment-mediator confounders (measured either at baseline or after treatment).

We explored the mediated effects of altruism and management by estimating a Oaxaca-Blinder decomposition of the treatment effect using DML, where the structural parameters of interest include assignment to treatment, the mediators, and an interaction term between AHME treatment and the mediators.²⁵ The Oaxaca-Blinder decomposition consistently estimates the direct and indirect effects when both the mediators and treatment assignment are (conditionally) exogenous [34]²⁶. We use this linear model to explore how experimentally-induced variation plus additional exogeneity assumptions can produce credible measures of the mediating role of provider preferences and management for the case where we rule out confounding of treatment-outcome, treatment-mediators, and mediators-outcome [39]. The estimating equation is:

$$y_i = \beta_0 + \beta_1 D_i + \beta_2 M_i + \beta_3 D_i \times M_i + g(X_i) + \varepsilon_i \quad (1.5)$$

Where D_i indicates assignment to treatment, M_i is an indicator variable that represent provider preferences and management. For provider preferences, $M_i = 1$ if the SP saw a provider falling into the groups defined by $\alpha \geq 0.75$, $\alpha \geq 0.9$. For management, M_i represents whether the provider was located in a clinic ranking above the median management score. We estimated the mediated proportions of provider preferences and management post-hoc by computing non-linear combinations of parameters. Specifically, mediated proportions were computed by dividing the coefficient of the interaction between AHME and the mediator by the total effect of AHME. Estimation was performed with the Stata command *xporegress*, and standard errors of the mediated proportions were derived via the delta method using the command *nlcom*.

$$\tau(d) = \beta_1 + \beta_3 \quad (1.6)$$

and

$$\delta(d) = \frac{\beta_3}{\beta_1 + \beta_3} \quad (1.7)$$

Mediated proportions of altruism and management

Table 1.14 presents the results of the mediation analysis using DML for six outcomes: Correct case management, price of visit, number of clients, revenues, profits, and unit cost of non-labor. Columns 1-3 show the results of single-mediator models, while models in columns 4 and 5 include both mediators.

Our results show support a mediation story. For correct case management, price of visit, number of clients, and revenues, the treatment effect of AHME is mediated by changes

²⁵Our relatively small sample size prevented us from applying doubly robust methods, while the IPTW approach was less efficient than the linear model, as extreme weights at the tails of the propensity score distribution increased the variance and decreased the balance between covariates.

²⁶Our model imposes the expectation of the outcome to be linear in the mediators.

in provider preferences, namely towards profit-maximization and away from altruism. For example, between 68% and 89% of the treatment effect of AHME on correct case management is driven by changes in the α parameter (columns 4 and 5), and these proportions are statistically significant at the 1% level. Similar patterns were found for price of visit, number of clients, and revenues. In the case of profits and unit cost of non-labor, provider preferences do not seem to play a role in explaining the treatment effect, but management does. For profits (columns 4 and 5), between 50% and 76% of the treatment effect is driven by changes in management adoption (significant at the 10% level), whereas for unit cost of non-labor, the model is consistent with full mediation as changes in management explain between 86% and 117% of the overall treatment effect (statistically significant at the 5% level).²⁷

6 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, SPs which are the gold standard for measuring provider practice, 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 health-care providers. Our rich set of baseline characteristics allowed us to decompose the mediating role of management and providers' social preferences on measures of clinical decision-making and firm profitability.

The first and foremost result of the evaluation is that a large-scale management consulting intervention in private for-profit health care clinics was successfully implemented as planned. While about a quarter of the treatment clinics did not participate in the whole intervention period, the rest had AHME interventions in place for at least 24 months before the endline evaluation. Hence, any null or negative results are most likely due to design issues as opposed to weak implementation problems.

Our study presents causal evidence of the relationship between management and firm performance in primary health care. We also find that 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 exhibited higher profit-maximizing preferences as a result of the program, and interactions with such providers showed significantly lower levels of correct case management including minimally essential actions. Moreover, these reductions were accompanied by significantly higher prices. Importantly, we also show

²⁷Mediated proportions greater than 100% are the result of "supressors" or negative mediated proportions in other mediator variables.

that provider preferences (but not management) mediate the treatment effects on process quality and price of visit.

Since we found that utilization, revenues, and prices increase due to the program without any changes to the number of clients in the waiting room, the amount of time a client spent with a provider, or staffing levels, we believe these increases were absorbed by excess capacity. Our data and causal mediation analysis is consistent with the notion of management as a technology, as it raises total factor productivity and increases firm profitability [13].

Our findings suggest that distributive preferences among private providers, particularly profit-maximization, play a significant role in the provision of quality and affordable care. We find that AHME changed the incentives of providers (that is, made them more profit-driven) without changing financial incentives, staff composition, or payroll expenses. 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 social preferences 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 processes and quality, as well as an awareness of how expenditures on costly 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 [5]; 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 consider is changing norms, such as an emphasis on fair-mindedness or altruistic preferences.

Our results do not mean that management training does not work in primary health care. The lack of skills and low levels of adoption of management techniques remain, and the emergence of public insurance and capitation schemes highlights the role of effective management, that spans from interacting with payers and regulators to overseeing the day-to-day operations of the clinic. As these insurance schemes operate under fee-for-service or capitation and reimbursements are time-discounted, the need for cost control and efficiency increases.

Management does matter — as posited by Bloom, incentivizing employees lies at the core of good management in health care. Despite of this, the emphasis is rarely on the employees but rather on the patients, while it's not the patients the physician leader manages [70]. Yet the role of the manager and the physician are in conflict because their intrinsic values are different. This is why many physicians do not like managerial roles, because they do not see management as consistent with their altruistic mission. Therefore, economic theory predicts managerial tasks will either be delegated to another person (which is costly) or eventually internalized by the physician leader.

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.

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). We obtained informed consent from all respondents of clinic, household, and exit interview surveys. Similar to other SP studies with similar designs and embedded in an intervention [56, 55], 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.

Figures and Tables

Figure 1: Sample and Experimental Design

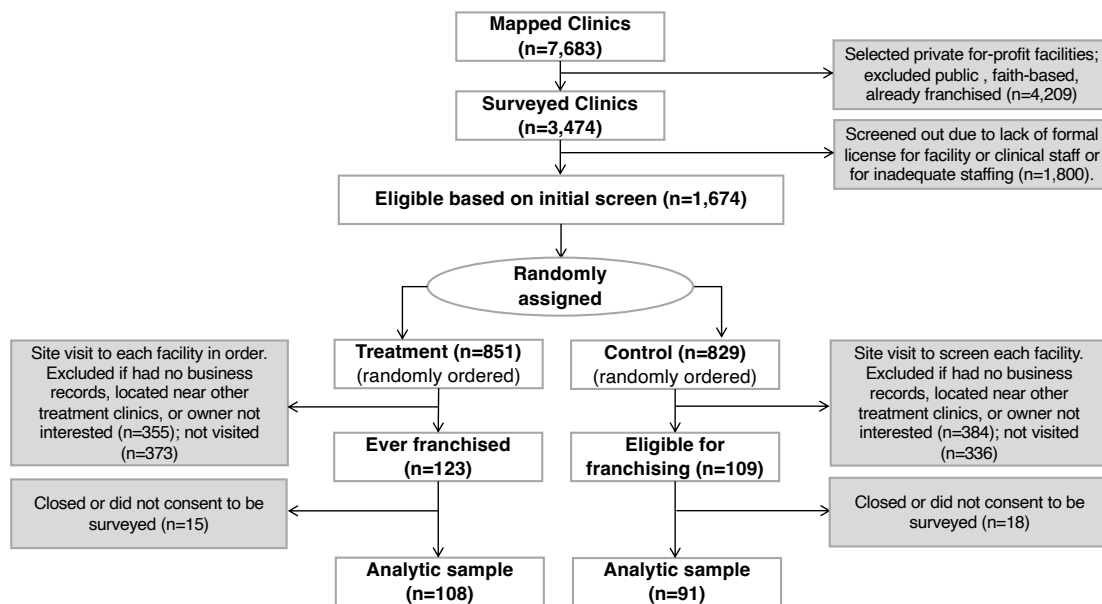


Figure 2: Evaluation Sample Locations

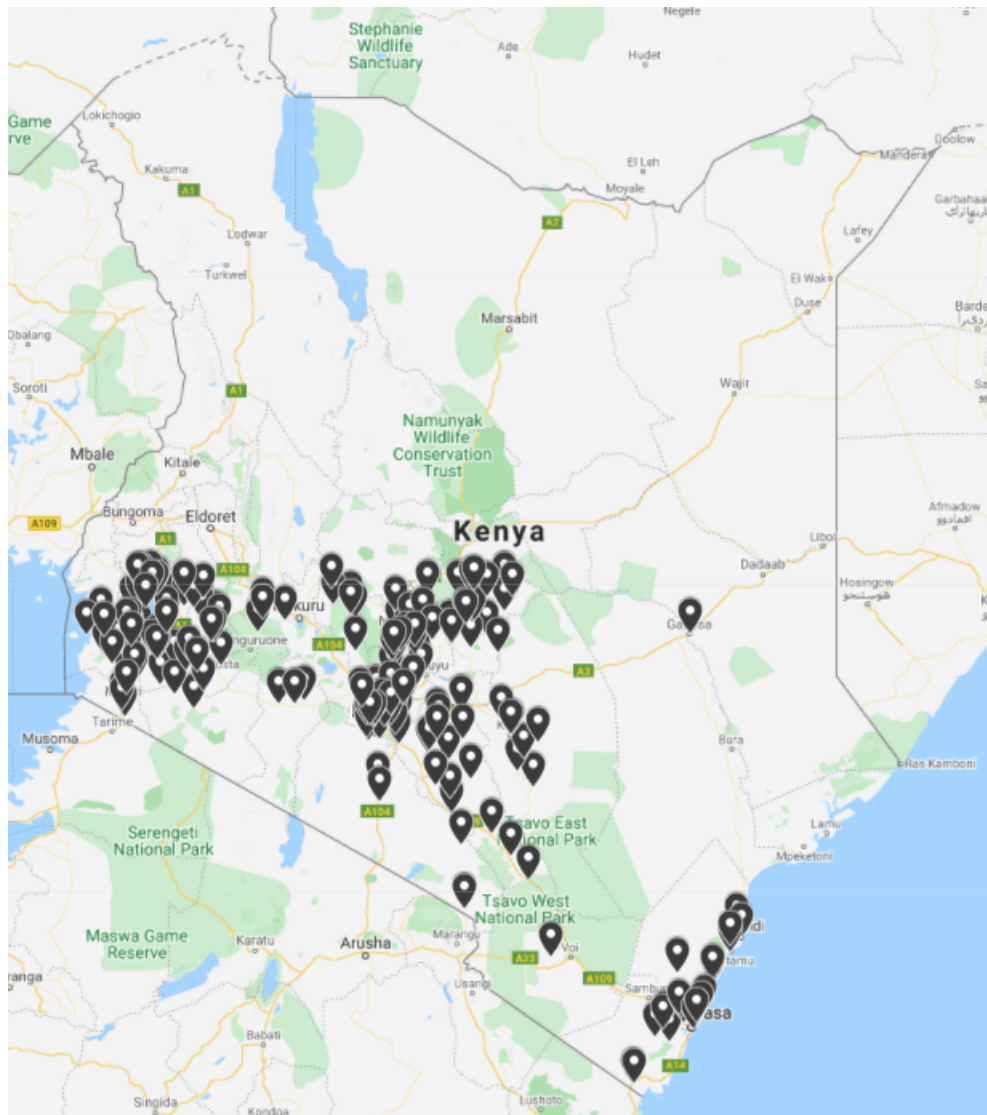


Figure 3: Treatment Exposure Time Distribution

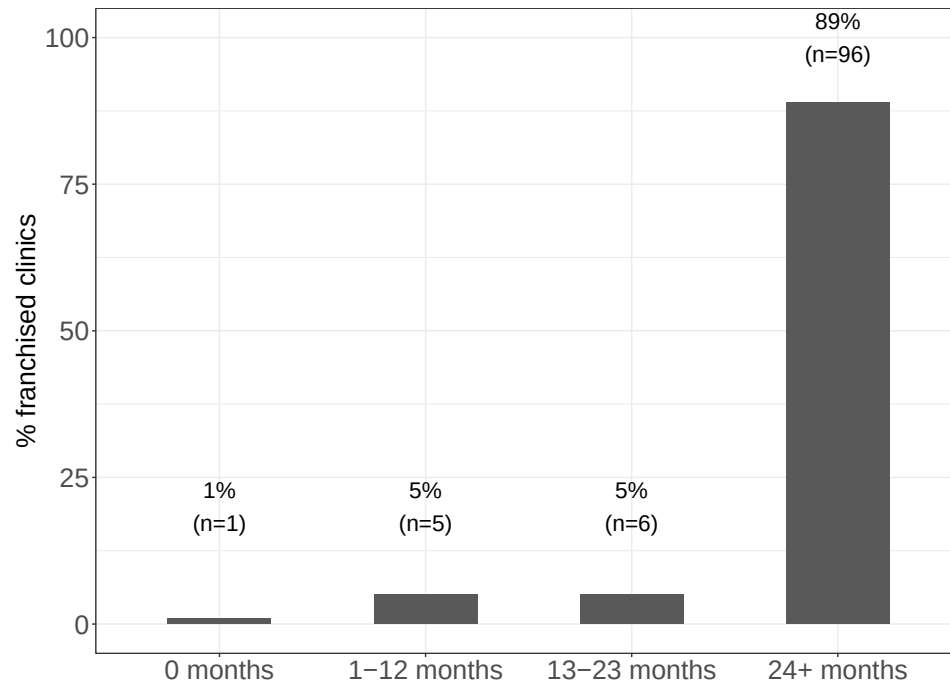


Table 1.1: Treatment Effects on Management Practices

	Operations management index (z-score)	Financial operations index (z-score)	Performance monitoring index (z-score)	Target setting index (z-score)	People management index (z-score)	Overall management index (z-score)
AHME treatment						
Coefficient	0.349	0.471	0.143	0.253	0.295	0.386
Standard Error	(0.148)	(0.161)	(0.148)	(0.170)	(0.147)	(0.157)
p -value	[0.009]	[0.002]	[0.167]	[0.069]	[0.022]	[0.007]
Adjusted p -value	<0.019>	<0.011>	<0.072>	<0.039>	<0.024>	-
Mean of control	0.00	0.00	0.00	0.00	0.00	0.00
Observations	191	191	191	190	191	191

Notes: Each column is a separate model estimated by Double Machine Learning (DML). Standard errors in parentheses; one-sided p -values are in brackets ($H_0 : \beta \leq 0$); Benjamini, Krieger, and Yekutieli (2006) sharpened q -vals in angled brackets.

Figure 4: Cumulative Distribution Function (CDF) of Overall Management (z-score)

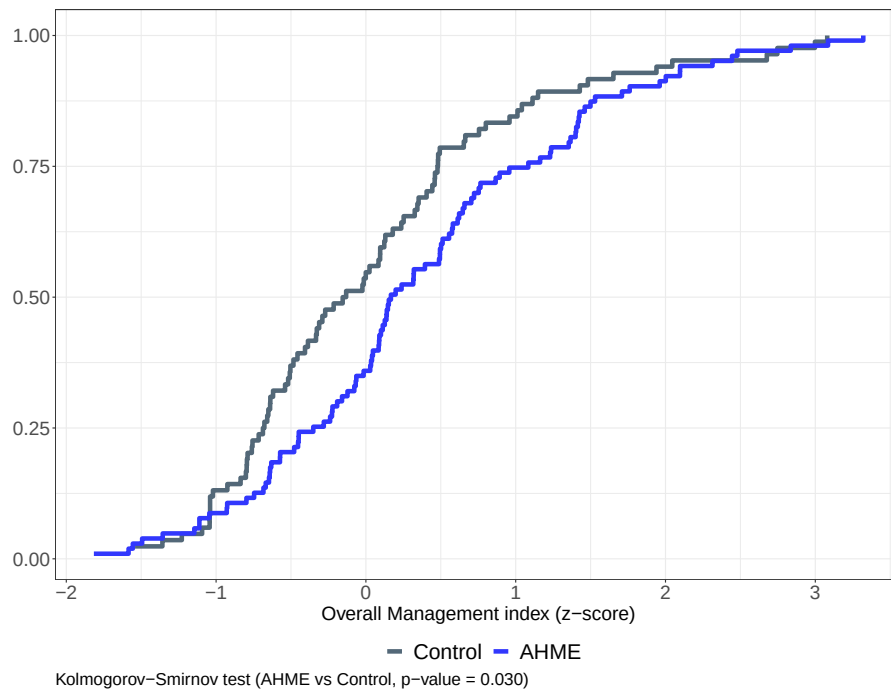


Table 1.2: Treatment Effects on Altruism

	(1) $\alpha \geq 0.50$	(2) $\alpha \geq 0.75$	(3) $\alpha \geq 0.90$
AHME treatment			
Coefficient	0.037	0.127	0.094
Standard Error	(0.057)	(0.065)	(0.056)
<i>p</i> -value	[0.261]	[0.025]	[0.046]
Adjusted <i>p</i> -value	<0.096>	<0.075>	<0.075>
Mean of control group	0.74	0.32	0.17
Observations	242	242	242
Case controls	Yes	Yes	Yes
Provider controls	Yes	Yes	Yes
Facility controls	Yes	Yes	Yes

Notes: Notes: Each column is a separate model estimated by Double Machine Learning (DML) on dichotomized versions of α at 0.5, 0.75, and 0.90. Cluster-robust standard errors are provided in parentheses; one-sided *p*-values in brackets; Benjamini Krieger Yekutieli (2006) sharpened *q*-vals in angled brackets.

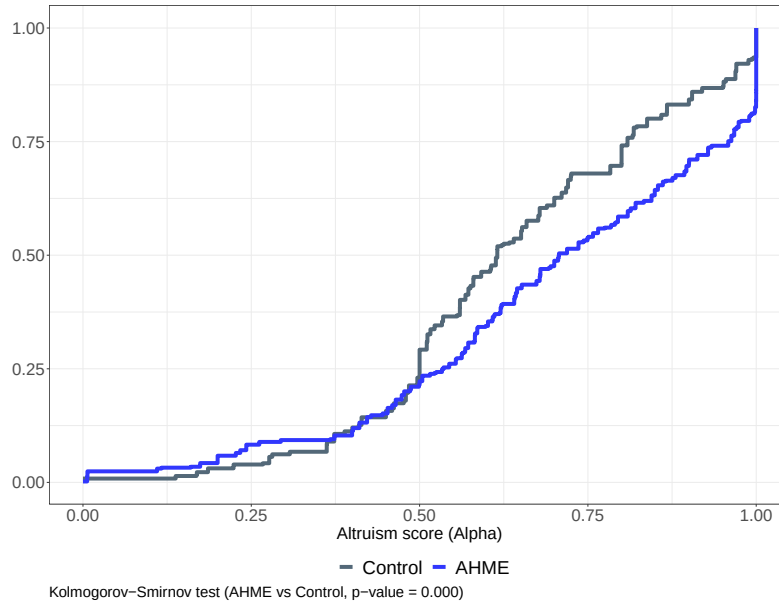
Figure 5: Cumulative Distribution Function (CDF) of profit-oriented preferences α 

Table 1.3: Quality of Care Conceptual Framework

	Structure	Process	Outcomes
	Attributes of material resources, human resources, and organizational structure in the setting where care occurs	Patients seeking and receiving care, as well as providers diagnosing and treating	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

Notes: Adapted from [24, 28, 27].

Table 1.4: Standardized Patient and Vignette Case Scenarios

Case	SP	Vignette	Case description	SP opening statement	Correct case management
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.”	Gave or advised on ORS, or Referred or asked to return
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.”	Asked FP method history, and Asked obstetric history, and Ruled out pregnancy, and Asked preferred FP method
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.”	Treated with inhaler or bronchodilator (e.g. salbutamol, cetirizine, prednisolone)
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.”	Ordered malaria rapid diagnostic test (mRDT) or malaria microscopy

Table 1.5: Treatment Effects on Structural Quality

	(1) Interpersonal structural quality (z-score)	(2) Technical structural quality (z-score)	(3) Overall structural quality (z-score)
AHME treatment			
Coefficient	0.281	0.150	0.521
Standard Error	(0.145)	(0.160)	(0.142)
p-value	[0.026]	[0.173]	[0.000]
Adjusted p-value	<0.060>	<0.060>	-
Mean of control	0.00	0.00	0.00
Observations	191	191	191

Notes: Each column is a separate model of difference in means estimated by OLS. Standard errors in parentheses; one-sided p -values are in brackets ($H_0 : \beta \leq 0$); Benjamini Krieger Yekutieli (2006) sharpened q -vals in angled brackets.

Figure 6: Cumulative Distribution Function (CDF) of Overall Structural Quality (z-score)

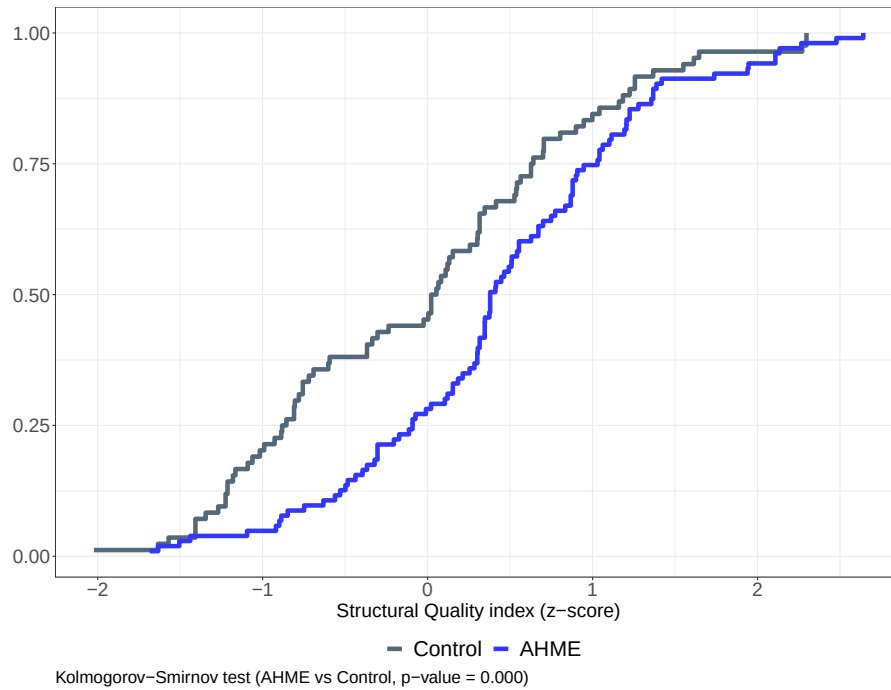


Table 1.6: Treatment Effects on Knowledge of Correct Case Management

	(1) Diarrhea case	(2) Family planning case	(3) Malaria case	(4) Knowledge index
AHME treatment				
Coefficient	-0.007	0.026	-0.015	-0.039
Standard Error	(0.040)	(0.047)	(0.016)	(0.129)
p-value	[0.569]	[0.290]	[0.822]	[0.617]
Adjusted p-value	<1.000>	<1.000>	<1.000>	-
Mean control group	0.909	0.113	0.990	0.000
Observations	240	237	239	241

Notes: Each column represents a different model estimated by Double Machine Learning (DML). Standard errors in parentheses; one-sided p -values are in brackets; Benjamini Krieger Yekutieli (2006) sharpened q -vals in angled brackets. Overall knowledge index computed as an Anderson z -score.

Table 1.7: Treatment Effects on Investment

	(1) Investment diagnostic and lab equipment	(2) Investment medical equipment	(3) Total investment
AHME treatment			
Coefficient	-0.203	0.442	0.421
Standard Error	(0.308)	(0.289)	(0.317)
p -value	[0.744]	[0.063]	[0.092]
Adjusted p -value	<0.593>	<0.145>	-
Mean of control	75	321	2529
Observations	191	191	191

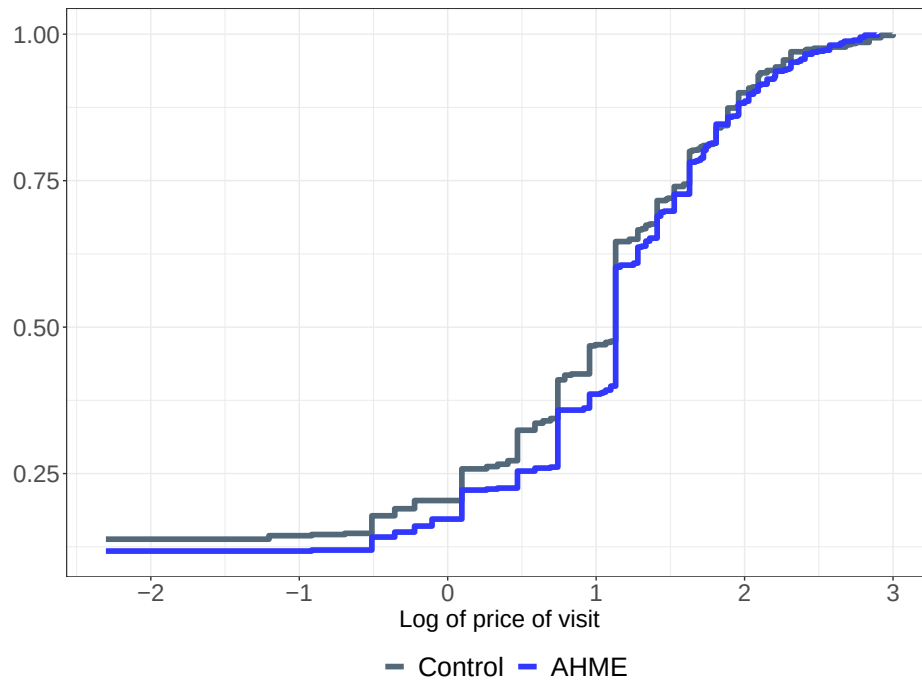
Notes: Each column is a separate model estimated with Double Machine Learning. Reported coefficients are semi-elasticities computed according to [9] since the dependent variable was transformed with the inverse hyperbolic sine (IHST) to deal with zeros. Standard errors in parentheses; one-sided p -values are in brackets ($H_0 : \beta \leq 0$); Benjamini Krieger Yekutieli (2006) sharpened q -vals in angled brackets. Mean of control group expressed in USD.

Table 1.8: Treatment Effects on Correct Case Management and Price

	(1)	(2)	(3)	(4)	(5)	(6)
	Correct case management					Price
	Asthma	Diarrhea	Family Planning	Malaria	Pooled	Price of visit (ln)
AHME treatment						
Coefficient	-0.152	-0.049	-0.018	-0.105	-0.078	0.269
Standard Error	(0.076)	(0.050)	(0.071)	(0.043)	(0.022)	(0.106)
<i>p</i> -value	[0.046]	[0.330]	[0.803]	[0.014]	[0.000]	[0.032]
Adjusted <i>p</i> -value	<0.076>	<0.198>	<0.473>	<0.076>	-	<0.076>
Mean of control group	0.500	0.686	0.264	0.850	0.636	340
Observations	204	400	199	392	1195	1195

Notes: Each column is a separate model estimated by Double Machine Learning (DML). Standard errors in parentheses, one-sided *p*-values in brackets; Benjamini Krieger Yekutieli (2006) sharpened *q*-vals in angled brackets. Mean of control group in model (6) expressed in Kenyan shillings (100 KSH $\sim$ 1 USD).

Figure 7: Cumulative Distribution Function (CDF) of Log of Price of Visit



Kolmogorov-Smirnov test (AHME vs Control, *p*-value = 0.040)

Table 1.9: Treatment Effects on Patient Satisfaction

	(1) 5-item Amenities index (z-score)	(2) Providers spent sufficient time (=1)	(3) Completely trust provider's medical treatment decision (=1)
(a) Standardized Patients			
AHME treatment			
Coefficient	0.102	0.025	-0.025
Standard Error	(0.064)	(0.032)	(0.026)
-value	[0.055]	[0.218]	[0.827]
Adjusted <i>p</i> -value	<0.199>	<0.279>	<0.486>
Mean of control group	0.000	0.498	0.782
Observations	1086	1086	1076
(b) Exit Interviews			
AHME treatment			
Coefficient	0.057	-0.031	0.029
Standard Error	(0.057)	(0.026)	(0.017)
-value	[0.158]	[0.885]	[0.047]
Adjusted <i>p</i> -value	<0.188>	<0.419>	<0.165>
Mean control group	0.000	0.544	0.888
Observations	1539	1531	1528

Notes: Each column is a separate model estimated by Double Machine Learning (DML). Estimated effects on client's perception and satisfaction with care using standardized patient (SP) and exit interviews (EI) data in the top and bottom panels, respectively. Cluster-robust standard errors in parentheses; one-sided *p*-values are in brackets ($H_0 : \beta \leq 0$); Benjamini Krieger Yekutieli (2006) sharpened *q*-vals in angled brackets. Models in the upper panel contain SP and case fixed effects (0-1 indicators for each SP experiment: demanding, unmarried, poor).

Table 1.10: Treatment Effects on Household Perceptions of the Clinic

	(1)	(2)
	Clinic quality rating (low)–10 (high) (<i>z</i> -score)	Clinic quality rating compared to its competitors (<i>z</i> -score)
AHME treatment		
Coefficient	0.021	0.041
Standard Error	(0.043)	(0.042)
<i>p</i> -value	[0.620]	[0.326]
Adjusted <i>p</i> -value	<1.000>	<1.000>
Mean of control group	0.000	0.000
Observations	2582	2582
Households	186	186

Notes: Each column is a separate model estimated by multivariate OLS on household survey data. Cluster-robust standard errors in parentheses; two-sided *p*-values are in brackets; Benjamini Krieger Yekutieli (2006) sharpened *q*-vals in angled brackets. Models control for the type of visit, where prenatal care is the reference category (regression coefficients not shown).

Table 1.11: Treatment Effects on Business Outcomes

	(1)	(2)	(3)
	Number of clients (ln)	Revenues (ln)	Profits (IHST)
AHME treatment			
Coefficient	0.293	0.379	0.948
Standard Error	(0.170)	(0.195)	(0.541)
p -value	[0.042]	[0.026]	[0.040]
Adjusted p -value	<0.045>	<0.045>	<0.045>
Mean of control group	385	3,340	167
Observations	187	184	180

Notes: Each column is a separate model estimated by Double Machine Learning (DML). Coefficients reported are semi-elasticities (computed according to [9] for model 3). Standard errors in parentheses; one-sided p -values are in brackets ($H_0 : \beta \leq 0$); Benjamini Krieger Yekutieli (2006) sharpened q -vals in angled brackets. Mean of the control group expressed in its natural units (clients for model 1, and USD for models 2 and 3).

Figure 8: Cumulative Distribution Function (CDF) of Number of Clients (ln)

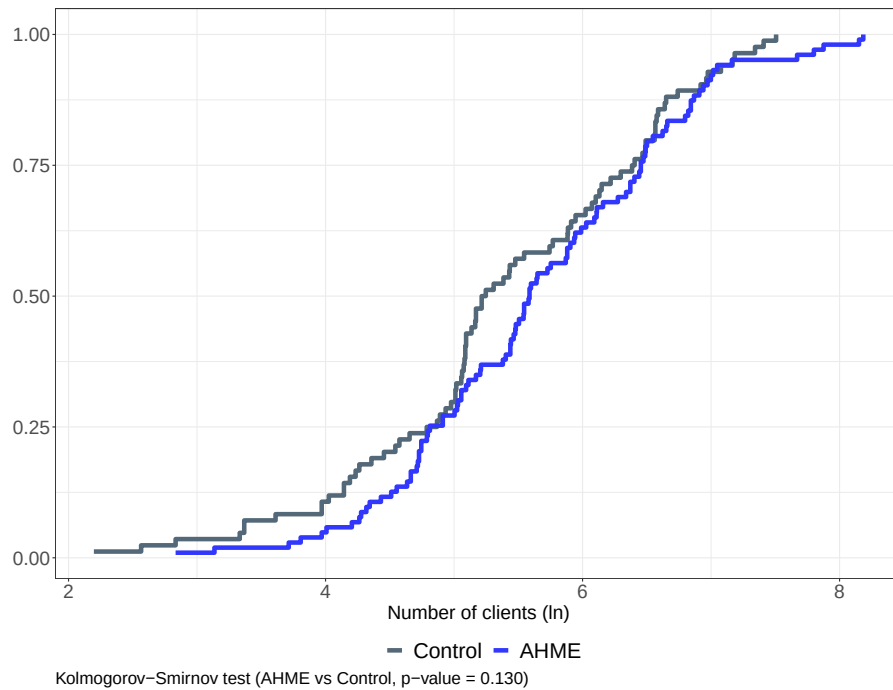


Figure 9: Cumulative Distribution Function (CDF) of Revenues (ln)

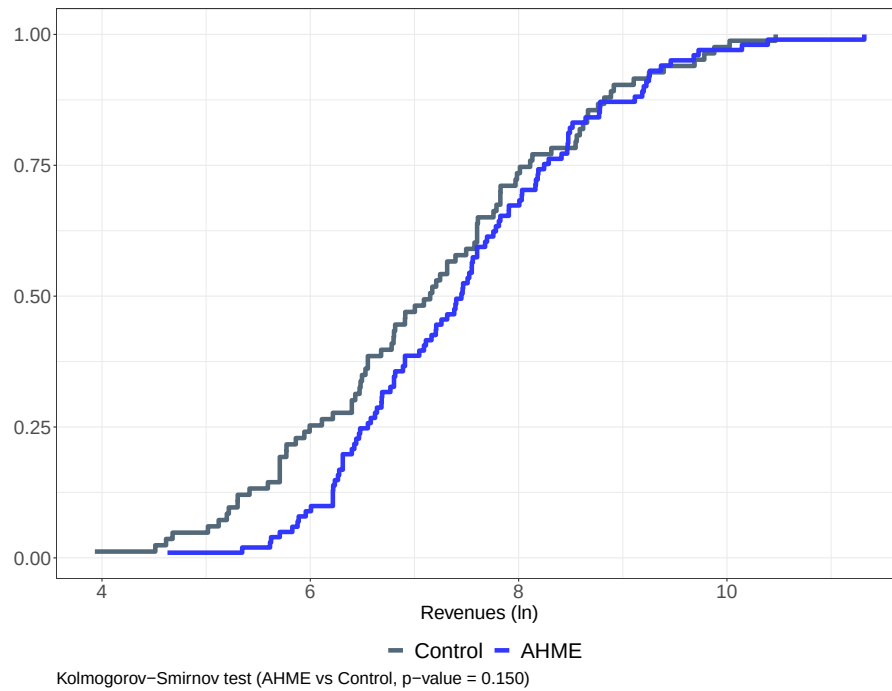


Figure 10: Cumulative Distribution Function (CDF) of Profits (IHST)

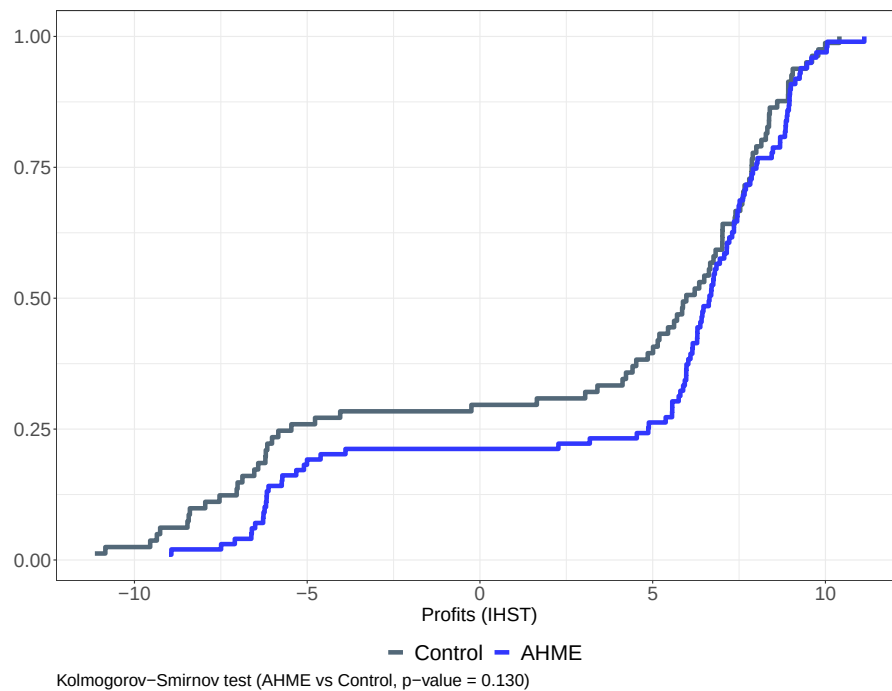


Table 1.12: Treatment Effects on Demand Generation Activities

	(1)	(2)	(3)	(4)
	Facility is NHIF empaneled (=1)	Facility performed marketing last year (=1)	Count of marketing activities performed	Facility has community health volunteers (=1)
AHME treatment				
Coefficient	0.156	0.374	0.975	0.293
Standard Error	(0.063)	(0.069)	(0.197)	(0.069)
p -value	[0.007]	[0.000]	[0.000]	[0.000]
Adjusted p -value	<0.002>	<0.001>	<0.001>	<0.001>
Mean of control	0.21	0.35	0.57	0.20
Observations	191	191	191	191

Notes: Each column is a separate model of difference in means estimated by Double Machine Learning (DML). Standard errors in parentheses; one-sided p -values are in brackets ($H_0 : \beta \leq 0$); Benjamini Krieger Yekutieli (2006) sharpened q -vals in angled brackets. NHIF stands for National Hospital Insurance Fund.

Table 1.13: Treatment Effects on Efficiency

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Unit cost	Unit cost clinical staff	Unit cost non-labor	Clients per clinical provider (ln)	Count of stockouts (last year)	Price of lab inputs (ln)	Price of drugs (ln)
AHME treatment							
Coefficient	-4.185	-0.218	-3.920	0.251	-0.140	-0.173	0.042
Standard Error	(2.548)	(0.663)	(2.481)	(0.170)	(0.356)	(0.053)	(0.027)
<i>p</i> -value	[0.050]	[0.371]	[0.057]	[0.070]	[0.347]	[0.001]	[0.936]
Adjusted <i>p</i> -value	<0.118>	<0.228>	<0.118>	<0.118>	<0.228>	<0.004>	<0.670>
Mean of control	10.4	2.7	6.9	210	1.20	3.695	24.77
Observations	182	179	175	174	191	18,489	6,483

Notes: Each column is a separate model of difference in means estimated by Double Machine Learning (DML). Estimated effects expressed as semi-elasticities for models (1)-(4) and (6). Standard errors in parentheses; one-sided *p*-values in brackets ($H_0 : \beta \leq 0$); Benjamini Krieger Yekutieli (2006) sharpened *q*-vals in angled brackets. Mean of control group expressed in USD for models (1),(3),(6), and (7); number of clients in column (4).

Figure 11: Cumulative Distribution Function (CDF) of Unit cost (ln)

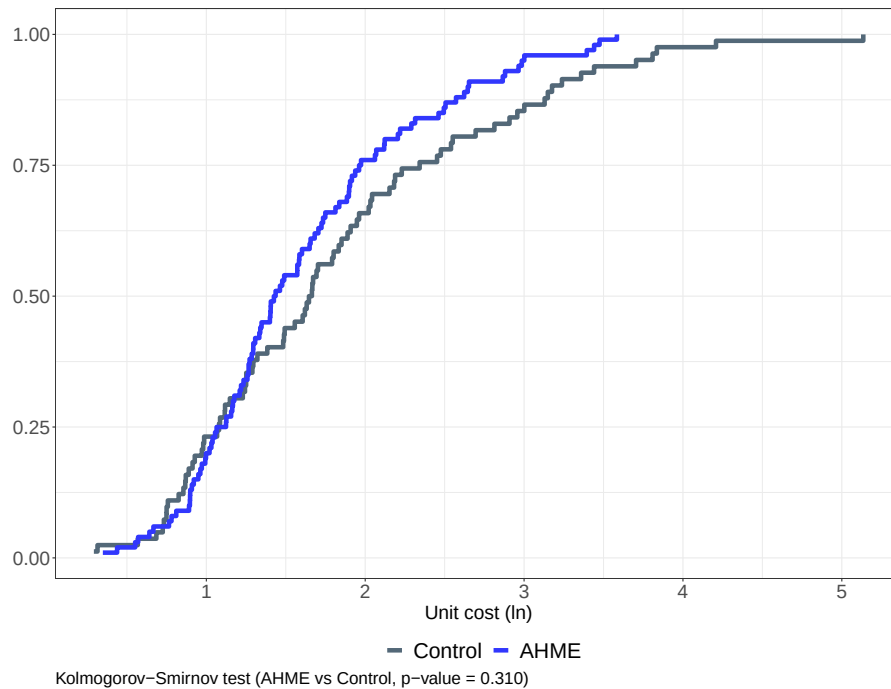


Table 1.14: Mediation Analysis of Altruism and Management

	(1) Alpha ≥ 0.75	(2) Alpha ≥ 0.90	(3) Mgmt $\geq p(50)$	(4) Alpha ≥ 0.75 Mgmt $\geq p(50)$	(5) Alpha ≥ 0.90 Mgmt $\geq p(50)$
Correct case management (n=641)					
Mediated proportion (Alpha)	1.054 (0.255) [0.000]	0.794 (0.267) [0.001]		0.890 (0.258) [0.000]	0.680 (0.268) [0.006]
Mediated proportion (Mgmt)			0.401 (0.642) [0.266]	0.212 (0.303) [0.242]	0.211 (0.335) [0.265]
Price of visit (ln, n=638)					
Mediated proportion (Alpha)	0.724 (0.337) [0.016]	0.868 (0.152) [0.000]		0.695 (0.391) [0.038]	0.855 (0.198) [0.000]
Mediated proportion (Mgmt)			-0.107 (1.014) [0.542]	0.098 (0.470) [0.418]	0.064 (0.248) [0.398]
Number of clients (ln, n=158)					
Mediated proportion (Alpha)	0.666 (0.381) [0.040]	0.504 (0.526) [0.169]		0.817 (0.680) [0.115]	0.442 (1.071) [0.340]
Mediated proportion (Mgmt)			-0.847 (3.153) [0.606]	-0.423 (1.053) [0.656]	-0.480 (1.582) [0.619]
Revenues (ln, n=156)					
Mediated proportion (Alpha)	0.888 (0.426) [0.019]	0.659 (0.451) [0.072]		0.850 (0.581) [0.072]	0.666 (0.451) [0.070]
Mediated proportion (Mgmt)			0.332 (1.487) [0.412]	0.124 (0.671) [0.427]	0.169 (0.538) [0.376]
Profits (IHST, n=155)					
Mediated proportion (Alpha)	0.261 (0.872) [0.382]	0.528 (0.616) [0.196]		0.112 (0.565) [0.422]	0.443 (0.372) [0.117]
Mediated proportion (Mgmt)			0.774 (0.515) [0.067]	0.757 (0.506) [0.068]	0.508 (0.391) [0.097]
Unit cost of non-labor (n=152)					
Mediated proportion (Alpha)	-1.914 (4.755) [0.656]	0.081 (1.048) [0.469]		-0.445 (0.813) [0.708]	0.033 (0.488) [0.473]
Mediated proportion (Mgmt)			0.883 (0.306) [0.002]	1.174 (0.550) [0.016]	0.862 (0.417) [0.019]

Notes: Each panel within the table presents the proportion of the overall treatment effect that is explained by altruism and management for seven outcomes. Proportions were obtained from a Oaxaca-Blinder decomposition of the treatment effect using a partially linear model, where controls were selected and nuisance functions estimated using Double/Debiased Machine Learning methods. Standard errors in parentheses, one-sided p -values in brackets ($H_0 : \beta \leq 0$).

Bibliography

- [1] George A Akerlof and Rachel E Kranton. “Economics and identity”. In: *The quarterly journal of economics* 115.3 (2000), pp. 715–753.
- [2] Michael L Anderson. “Multiple inference and gender differences in the effects of early intervention: A reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects”. In: *Journal of the American statistical Association* 103.484 (2008), pp. 1481–1495.
- [3] Stephen J Anderson, Christy Marie Lazicky, and Bilal Husnain Zia. *Measuring the Unmeasured: Combining Technology and Behavioral Insights to Improve Measurement of Business Outcomes*. The World Bank, 2019.
- [4] James Andreoni and John Miller. “Giving according to GARP: An experimental test of the consistency of preferences for altruism”. In: *Econometrica* 70.2 (2002), pp. 737–753.
- [5] Kenneth J Arrow. “Uncertainty and the welfare economics of medical care”. In: *The American economic review* 53.5 (1963), pp. 941–973.
- [6] Uttara Balakrishnan, Johannes Haushofer, and Pamela Jakiela. “How soon is now? Evidence of present bias from convex time budget experiments”. In: *Experimental Economics* 23.2 (2020), pp. 294–321.
- [7] Jeff Barnes et al. *Private health sector assessment in Kenya*. The World Bank, 2010.
- [8] Reuben M Baron and David A Kenny. “The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations.” In: *Journal of personality and social psychology* 51.6 (1986), p. 1173.
- [9] Marc F Bellemare and Casey J Wichman. “Elasticities and the inverse hyperbolic sine transformation”. In: *Oxford Bulletin of Economics and Statistics* 82.1 (2020), pp. 50–61.
- [10] Roland Bénabou and Jean Tirole. “Individual and corporate social responsibility”. In: *Economica* 77.305 (2010), pp. 1–19.
- [11] B Douglas Bernheim et al. “A theory of chosen preferences”. In: *American Economic Review* 111.2 (2021), pp. 720–54.

- [12] Nicholas Bloom, Raffaella Sadun, and John Van Reenen. “Does management matter in healthcare”. In: *Boston, MA: Center for Economic Performance and Harvard Business School* (2014).
- [13] Nicholas Bloom et al. *Do management interventions last? evidence from India*. The World Bank, 2018.
- [14] Nicholas Bloom et al. “Does management matter? Evidence from India”. In: *The Quarterly Journal of Economics* 128.1 (2013), pp. 1–51.
- [15] Miriam Bruhn, Dean Karlan, and Antoinette Schoar. “The impact of consulting services on small and medium enterprises: Evidence from a randomized trial in Mexico”. In: *Journal of Political Economy* 126.2 (2018), pp. 635–687.
- [16] John B Burbidge, Lonnie Magee, and A Leslie Robb. “Alternative transformations to handle extreme values of the dependent variable”. In: *Journal of the American Statistical Association* 83.401 (1988), pp. 123–127.
- [17] Lisa Cameron and Manisha Shah. “Risk-taking behavior in the wake of natural disasters”. In: *Journal of Human Resources* 50.2 (2015), pp. 484–515.
- [18] Victor Chernozhukov et al. *Double/debiased machine learning for treatment and structural parameters*. 2018.
- [19] Janet Currie, Wanchuan Lin, and Wei Zhang. “Patient knowledge and antibiotic abuse: Evidence from an audit study in China”. In: *Journal of health economics* 30.5 (2011), pp. 933–949.
- [20] Benjamin Daniels et al. “Use of standardised patients to assess quality of healthcare in Nairobi, Kenya: a pilot, cross-sectional study with international comparisons”. In: *BMJ global health* 2.2 (2017), e000333.
- [21] Jishnu Das et al. “In urban and rural India, a standardized patient study showed low levels of provider training and huge quality gaps”. In: *Health affairs* 31.12 (2012), pp. 2774–2784.
- [22] Jishnu Das et al. “Use of standardised patients to assess quality of tuberculosis care: a pilot, cross-sectional study”. In: *The Lancet infectious diseases* 15.11 (2015), pp. 1305–1313.
- [23] Laura Di Giorgio et al. “Analysis of clinical knowledge, absenteeism and availability of resources for maternal and child health: a cross-sectional quality of care study in 10 African countries”. In: *BMJ global health* 5.12 (2020), e003377.
- [24] Avedis Donabedian. “Evaluating the quality of medical care”. In: *The Milbank memorial fund quarterly* 44.3 (1966), pp. 166–206.
- [25] Avedis Donabedian. “Promoting quality through evaluating the process of patient care”. In: *Medical care* 6.3 (1968), pp. 181–202.
- [26] Avedis Donabedian. “The methods and findings of quality assessment and monitoring: an illustrated analysis”. In: *Journal for Healthcare Quality* 7.3 (1985), p. 15.

- [27] Avedis Donabedian. “The quality of care: how can it be assessed?” In: *Jama* 260.12 (1988), pp. 1743–1748.
- [28] Avedis Donabedian. “The quality of medical care”. In: *Science* 200.4344 (1978), pp. 856–864.
- [29] David Dranove and Mark A Satterthwaite. “Monopolistic competition when price and quality are imperfectly observable”. In: *The RAND Journal of Economics* (1992), pp. 518–534.
- [30] Felipe A Dunsch et al. *Management, Supervision, and Health Care: A Field Experiment*. Tech. rep. National Bureau of Economic Research, 2017.
- [31] Tore Ellingsen and Magnus Johannesson. “Pride and prejudice: The human side of incentive theory”. In: *American economic review* 98.3 (2008), pp. 990–1008.
- [32] Timothy Evans, Margaret Whitehead, and Finn Diderichsen. *Challenging inequities in health: from ethics to action*. Oxford University Press, 2001.
- [33] Helmut Farbmacher et al. “Causal mediation analysis with double machine learning”. In: *arXiv preprint arXiv:2002.12710* (2020).
- [34] Ana Fernandes, Martin Huber, and Giannina Vaccaro. “Gender differences in wage expectations”. In: *Plos one* 16.6 (2021), e0250892.
- [35] Raymond Fisman, Pamela Jakiela, and Shachar Kariv. “Distributional preferences and political behavior”. In: *Journal of Public Economics* 155 (2017), pp. 1–10.
- [36] Anne Fitzpatrick and Katherine Tumlinson. “Strategies for optimal implementation of simulated clients for measuring quality of care in low-and middle-income countries”. In: *Global Health: Science and Practice* 5.1 (2017), pp. 108–114.
- [37] Sherman Folland, Allen C Goodman, Miron Stano, et al. *The economics of health and health care*. Vol. 6. Pearson Prentice Hall Upper Saddle River, NJ, 2007.
- [38] Paul J Gertler and Donald M Waldman. “Quality-adjusted cost functions and policy evaluation in the nursing home industry”. In: *Journal of Political Economy* 100.6 (1992), pp. 1232–1256.
- [39] James J Heckman and Rodrigo Pinto. “Econometric mediation analyses: Identifying the sources of treatment effects from experimentally estimated production technologies with unmeasured and mismeasured inputs”. In: *Econometric reviews* 34.1-2 (2015), pp. 6–31.
- [40] Martin Huber. “Causal pitfalls in the decomposition of wage gaps”. In: *Journal of Business & Economic Statistics* 33.2 (2015), pp. 179–191.
- [41] Martin Huber. “Mediation analysis”. In: *Handbook of Labor, Human Resources and Population Economics* (2020), pp. 1–38.
- [42] Jeremiah Hurley. “An overview of the normative economics of the health sector”. In: *Handbook of health economics*. Vol. 1. Elsevier, 2000, pp. 55–118.

- [43] Kosuke Imai, Luke Keele, and Dustin Tingley. “A general approach to causal mediation analysis.” In: *Psychological methods* 15.4 (2010), p. 309.
- [44] Kosuke Imai et al. “Unpacking the black box of causality: Learning about causal mechanisms from experimental and observational studies”. In: *American Political Science Review* 105.4 (2011), pp. 765–789.
- [45] Pamela Jakiela. “Equity vs. efficiency vs. self-interest: on the use of dictator games to measure distributional preferences”. In: *Experimental Economics* 16.2 (2013), pp. 208–221.
- [46] Dean Karlan and Martin Valdivia. “Teaching entrepreneurship: Impact of business training on microfinance clients and institutions”. In: *Review of Economics and statistics* 93.2 (2011), pp. 510–527.
- [47] Jacob S Kazungu and Edwine W Barasa. “Examining levels, distribution and correlates of health insurance coverage in Kenya”. In: *Tropical Medicine & International Health* 22.9 (2017), pp. 1175–1185.
- [48] Kenya Ministry of Health. *National Family Planning Guidelines for Service Providers*. Kenya Ministry of Health, Division of Reproductive Health, 2010.
- [49] Kenya Ministry of Health. *National Guidelines for Diagnosis, Treatment and Prevention of Malaria in Kenya*. Kenya Ministry of Health, 2010.
- [50] Kenya Ministry of Health. *Policy Guidelines for Management of Diarrhoea in Children below Five Years in Kenya*. Kenya Ministry of Health, 2014.
- [51] Kenya Ministry of Public Health and Sanitation. *Guidelines for Asthma Management in Kenya*. Kenya Ministry of Health, 2010.
- [52] Kenya President’s Malaria Initiative. *Kenya Malaria Operational Plan FY 2018*. President’s Malaria Initiative, 2018.
- [53] Jessica JC King et al. “How to do (or not to do)... using the standardized patient method to measure clinical quality of care in LMIC health facilities”. In: *Health policy and planning* 34.8 (2019), pp. 625–634.
- [54] Jonathan T Kolstad. “Information and quality when motivation is intrinsic: Evidence from surgeon report cards”. In: *American Economic Review* 103.7 (2013), pp. 2875–2910.
- [55] Ada Kwan et al. “Use of standardised patients for healthcare quality research in low- and middle-income countries”. In: *BMJ global health* 4.5 (2019), e001669.
- [56] Ada Kwan et al. “Variations in the quality of tuberculosis care in urban India: A cross-sectional, standardized patient study in two cities”. In: *PLoS medicine* 15.9 (2018), e1002653.
- [57] Jing Li, William H Dow, and Shachar Kariv. “Social preferences of future physicians”. In: *Proceedings of the National Academy of Sciences* 114.48 (2017), E10291–E10300.

- [58] Ulrike Malmendier. “Exposure, experience, and expertise: why personal histories matter in economics”. In: *Journal of the European Economic Association* (2021).
- [59] David McKenzie. “Small Business Training to Improve Management Practices in Developing Countries”. In: (2020).
- [60] David McKenzie and Christopher Woodruff. “What are we learning from business training and entrepreneurship evaluations around the developing world?” In: *The World Bank Research Observer* 29.1 (2014), pp. 48–82.
- [61] Kenya MoH. “Kenya Health Policy 2012 2030”. In: *Ministry of Medical Service: Ministry of Public Health and Sanitation* (2012).
- [62] Manoj Mohanan et al. “The know-do gap in quality of health care for childhood diarrhea and pneumonia in rural India”. In: *JAMA pediatrics* 169.4 (2015), pp. 349–357.
- [63] Joseph P Newhouse. “Toward a theory of nonprofit institutions: An economic model of a hospital”. In: *The American economic review* 60.1 (1970), pp. 64–74.
- [64] NMCP, KNBS, and ICF. *Kenya Malaria Indicator Survey 2015*.
- [65] John W Peabody et al. “Comparison of vignettes, standardized patients, and chart abstraction: a prospective validation study of 3 methods for measuring quality”. In: *Jama* 283.13 (2000), pp. 1715–1722.
- [66] James M Robins and Sander Greenland. “Identifiability and exchangeability for direct and indirect effects”. In: *Epidemiology* (1992), pp. 143–155.
- [67] Joseph P Romano and Michael Wolf. “Efficient computation of adjusted p-values for resampling-based stepdown multiple testing”. In: *Statistics & Probability Letters* 113 (2016), pp. 38–40.
- [68] Jeffrey D Sachs. “Achieving universal health coverage in low-income settings”. In: *The Lancet* 380.9845 (2012), pp. 944–947.
- [69] Sean Sylvia et al. “Survey using incognito standardized patients shows poor quality care in China’s rural clinics”. In: *Health policy and planning* 30.3 (2014), pp. 322–333.
- [70] J Michael Syptak. “Altruism in practice management: Caring for your staff”. In: *Family Practice Management* 5.9 (1998), p. 58.
- [71] Thomas C Tsai et al. “Hospital board and management practices are strongly related to hospital performance on clinical quality metrics”. In: *Health affairs* 34.8 (2015), pp. 1304–1311.
- [72] WHO. “Primary health care on the road to universal health coverage: 2019 global monitoring report”. In: (2021).

1 Appendix 1. 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 Foreign, Commonwealth & Development Office (FCDO, formerly 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, 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. **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.

2 Appendix 2. Supplement for Methods

Experimental Design

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 476 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 recruitment 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, 489 control clinics were identified for evaluation recruitment team visits to replicate franchise recruitment procedures used for treatment arm clinics. Of these 489 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.

Program Compliance

To understand the extent of intervention 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 both treatment and control clinics. 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.

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.

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. 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. For the analytic sample of process quality and health care outcomes, the SP exercise was conducted in 211 clinics belonging to the complete AHME evaluation clinic sample of 232. The other endline surveys (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).

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 A1 reports treatment- 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 managers in control clinics had a clinical degree); demographic characteristics (gender, age); and experience (tenure, positions at other clinics) (see table 1.B).

Table A1 reports balance 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 (total number of client visits), 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).

Table A1: 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.

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 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, 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.²⁸

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 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

²⁸Of 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.

Table A2: Balance of Clinic Manager Characteristics at Endline by AHME Treatment

	Control Group (n=91)	Treatment Group (n=108)	Difference <i>p</i> -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

Notes: Proportions reported, unless otherwise noted.

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 describe the sampling process of households attached to the clinic evaluation sample and describe the creation of the endline analytic sample. We then compare household characteristics collected at baseline to assess how the AHME household sample compared to a nationally representative sample of Kenyan households, and whether randomization created interchangeable groups in terms of their observable and unobservable characteristics.

Household Sample Recruitment from the Clinic Evaluation Sample

To identify households associated with one of the 232 evaluation clinics, we conducted exit interviews at baseline in 232 clinics in the AHME evaluation sample. We recruited 2127 households from 216 clinics. Eligible households could not be recruited at baseline in 8 treatment clinics and 9 control clinics for the following reasons: 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. 10 clinics were excluded from the clinic and quality of care analytic samples at endline. In total, we conducted 2,095 household interviews at baseline, associated with the remaining 206 clinics to comprise our target panel household sample. At endline, we completed 1,295 endline household surveys for 199 clinics. For 7 clinics, none of the associated 23 households could be contacted. We did not complete 747 panel endline surveys 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 reasons (n=110).

Description of AHME Household Sample

The AHME evaluation took place in seven provinces in Kenya: Central, Coast, Eastern, Nairobi, Nyanza Rift Valley and Western. We report how the households in the AHME analytic sample compared to a representative sample of households in these regions based on baseline survey data collected from AHME households during 2015 and 2016. Demographic and Health Surveys (DHS) fielded the Kenya Malaria Indicator Survey (MIS) in 2015 [64]. 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. The DHS data included a list of household assets that was amenable for creating a nationally representative standardized asset wealth score and nationally representative indicators of wealth quintiles. Since we harmonized AHME surveys with DHS and were fielded contemporaneously, the weights used to calculate the national wealth score could be applied to the AHME sample of households, and therefore have a direct comparison of asset wealth between the two samples. 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 households in the DHS sample, AHME households in the same province that utilized health services at AHME clinics in 2015 or 2016 were on average i) slightly more likely to be urban (AHME: 42%; DHS: 41%) and ii) to have more household members (median household members, AHME: 5; DHS: 3). Heads of AHME households were more likely to be male (AHME: 89%; DHS 64%), younger in terms of median age (AHME: 35 years; DHS 39), and more likely to have any education (AHME: 96%; DHS 87%). AHME households were 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%) as well as their own, non-shared, improved sanitation facility (AHME: 47%; DHS 28%). The distribution of household asset wealth among AHME and DHS households are similar. However, the AHME distribution of the asset ownership score is truncated at both tails, suggesting it has proportionally fewer

“very poor” and “very rich” households compared to representative households within the same geographical locations. We also found that the DHS households in the regions where AHME was implemented followed the same quintile distribution as the national population (20% in each quintile). Households comprising the endline AHME sample were less likely to be in the poorest quintile (AHME: 13%; DHS 20%) and wealthiest quintile (AHME: 10%; DHS 20%), and were more likely to have median levels of wealth.

In summary, households in the AHME household analytic sample at appeared to have higher levels of wealth, education, and improved infrastructure compared to a representative sample of households from the regions where AHME was implemented during the same period of 2015 and 2016. This evidence suggests that clinics in the AHME sample offered health care services to wealthier households at baseline, compared to an average household in the same province.

Baseline Balance of Household Sample

Within the AHME sample, we evaluated the balance of demographic, socioeconomic, and infrastructure indicators collected at baseline. Table 1.C 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 suggests that the household control arm provides an adequate counterfactual for the AHME treatment group, and therefore the use of the ITT estimate to obtain unbiased estimates of the impact of AHME on household-level outcomes.

Construction of Economic Outcomes

Self-reported economic outcomes in small firms in low- and middle-income countries are prone to substantial measurement error and non-response. Below, we describe the procedures used to construct three outcomes: labor costs, non-labor costs, and total number of clients.

1. *Labor 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 provider-level observations from the staff roster module of the clinic survey.
2. *Non-labor costs.* In four clinics, 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

Table A3: 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	
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	
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 (DHS) ^a	0.3	(0.07)	0.18	(0.09)	0.13	1.28
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	
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*	
Child in household had fever ^c	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	

Note: *** p<0.01, ** p<0.05, * p<0.1. Robust standard errors in parentheses, clustered by index clinic. Column (4) displays F-statistics for groups of baseline variables. All missing values in balance variables are treated as zero.

a. The wealth score is from the wealth index of the Demographic Health Survey (DHS).

b. Ratings on a 1-5 Likert scale, from 1 (worst/lowest) to 5 (best/highest) were recoded to binary indicator variables where 1 refers to ratings of 4 or 5 and 0 refers to ratings of 1, 2, or 3.

c. Prevalence and utilizations are from the 7 days before baseline survey. The wealth score is a 0-1 continuous variable, and here is normalized to a standard deviation of 1 within each group.

Table A4: 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)
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)

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. *Total number of clients.* 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 Appendix 3. Supplement for Results (ITT difference in means)

Table C1: Treatment Effects on Management Practices (ITT OLS)

	Operations management index (z-score)	Financial operations index (z-score)	Performance monitoring index (z-score)	Target setting index (z-score)	People management index (z-score)	Overall management index (z-score)
AHME treatment						
Coefficient	0.323	0.408	0.063	0.217	0.286	0.337
Standard Error	(0.141)	(0.155)	(0.149)	(0.168)	(0.141)	(0.151)
<i>p</i> -value	[0.012]	[0.004]	[0.337]	[0.099]	[0.022]	[0.013]
Adjusted <i>p</i> -value	<0.036>	<0.017>	<0.336>	<0.155>	<0.053>	-
Mean of control	0.00	0.00	0.00	0.00	0.00	0.00
Observations	191	191	191	190	191	191

Notes: Each column is a separate model of difference in means estimated by OLS. 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.

Table C2: Treatment Effects on Structural Quality (ITT OLS)

	(1) Interpersonal structural quality (z-score)	(2) Technical structural quality (z-score)	(3) Overall structural quality (z-score)
AHME treatment			
Coefficient	0.248	0.117	0.472
Standard Error	(0.141)	(0.157)	(0.137)
<i>p</i> -value	[0.041]	[0.229]	[0.000]
Adjusted <i>p</i> -value	<0.074>	<0.240>	<0.003>
Mean of control	0.00	0.00	0.00
Observations	191	191	191

Notes: Each column is a separate model of difference in means estimated by OLS. 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.

Table C3: Treatment Effects on Knowledge of Correct Case Management (ITT OLS)

	(1) Diarrhea case	(2) Family planning case	(3) Malaria case	(4) Knowledge index
AHME treatment				
Coefficient	-0.008	0.037	-0.018	-0.032
Standard Error	(0.039)	(0.045)	(0.019)	(0.138)
p-value	[0.414]	[0.790]	[0.164]	[0.407]
Adjusted p-value	<0.691>	<0.047>	<0.979>	-
Mean of control group	0.909	0.113	0.990	0.000
Observations	240	237	239	241

Notes: Each column represents a different model of difference in means estimated with multivariate OLS. Standard errors in parentheses; two-sided p -values are in brackets; adjusted p -values for multiple hypothesis testing (Romano & Wolf) in angled brackets. Overall knowledge index computed as an Anderson z -score.

Table C4: Treatment Effects on Investment (ITT OLS)

	(1) Investment diagnostic and lab equipment	(2) Investment medical equipment	(3) Total investment
AHME treatment			
Coefficient	-0.205	0.444	0.406
SE	(0.300)	(0.286)	(0.300)
p-value	[0.753]	[0.061]	[0.089]
Adjusted p-value	<0.747>	<0.099>	-
Mean of control group	75	321	2529
Observations	191	191	191

Notes: Each column is a separate model of difference in means estimated by OLS. Reported coefficients are semi-elasticities computed according to [9] since the dependent variable was transformed with the inverse hyperbolic sine (IHST) to deal with zeros. 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. Mean of control group expressed in USD.

Table C5: Treatment Effects on Patient Satisfaction (ITT OLS)

	(1) 5-item Amenities index (<i>z</i> -score)	(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)
<i>p</i>-value	[0.182]	[0.561]	[0.450]
Adjusted <i>p</i> -value	<0.224>	<0.482>	<0.767>
Mean of 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)
<i>p</i>-value	[0.306]	[0.204]	[0.163]
Adjusted <i>p</i> -value	<0.285>	<0.090>	<0.209>
Mean control group	0.000	0.544	0.888
Observations	1539	1531	1528

Notes: Each column is a separate model estimated by multivariate OLS. Estimated effects on client's perception and satisfaction with care using standardized patient (SP) and exit interviews (EI) data in the top and bottom panels, respectively. Cluster-robust 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 in the upper panel contain SP and case fixed effects (0-1 indicators for each SP experiment: demanding, unmarried, poor).

Table C6: Treatment Effects on Household Perceptions of the Clinic (ITT OLS)

	(1)	(2)
	Clinic quality rating (low)–10 (high) (<i>z</i> -score)	Clinic quality rating compared to its competitors (<i>z</i> -score)
AHME treatment		
Coefficient	0.039	0.050
Standard Error	(0.054)	(0.048)
<i>p</i> -value	[0.471]	[0.302]
Adjusted <i>p</i> -value	<0.493>	<0.493>
Mean of control group	0.000	0.000
Observations	2637	2637
Households	1139	1139
Clinics	192	192

Notes: Each column is a separate model estimated by multivariate OLS on household survey data. Cluster-robust standard errors in parentheses; two-sided *p*-values are in brackets; adjusted *p*-values for multiple hypothesis testing (Romano & Wolf) in angled brackets. Models control for the type of visit, where prenatal care is the reference category (regression coefficients not shown).

Table C7: Treatment Effects on Business Outcomes (ITT OLS)

	(1)	(2)	(3)
	Number of clients (ln)	Revenues (ln)	Profits (IHST)
AHME treatment			
Coefficient	0.260	0.339	0.898
SE	(0.162)	(0.195)	(0.532)
<i>p</i> -value	[0.056]	[0.042]	[0.047]
Adjusted <i>p</i> -value	<0.116>	<0.116>	<0.116>
Mean of control group	385	3,340	167
Observations	187	184	180

Notes: Each column is a separate model of difference in means estimated by OLS. Coefficients reported are semi-elasticities (computed according to [9] for model 3). 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. Mean of the control group expressed in its natural units (clients for model 1, and USD for models 2 and 3).

Table C8: Treatment Effects on Demand Generation Activities (ITT OLS)

	(1)	(2)	(3)	(4)
	Facility is NHIF empaneled (=1)	Facility performed marketing last year (=1)	Count of marketing activities performed	Facility has community health volunteers (=1)
AHME treatment				
Coefficient	0.166	0.373	0.954	0.291
SE	(0.066)	(0.067)	(0.201)	(0.067)
<i>p</i> -value	[0.007]	[0.000]	[0.000]	[0.000]
Adjusted <i>p</i> -value	<0.008>	<0.001>	<0.001>	<0.001>
Mean of control	0.21	0.35	0.57	0.20
Observations	191	191	191	191

Notes: Each column is a separate model of difference in means estimated by OLS. 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. NHIF stands for National Hospital Insurance Fund.

Table C9: Treatment Effects on Efficiency (ITT OLS)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Unit cost	Unit cost clinical staff	Unit cost non-labor	Clients per clinical provider (ln)	Count of stockouts (last year)	Price of lab inputs (ln)	Price of drugs (ln)
AHME treatment							
Coefficient	-4.619	-0.165	-4.192	0.248	-0.153	-0.172	0.044
SE	(2.243)	(0.646)	(2.175)	(0.160)	(0.335)	(0.111)	(0.043)
<i>p</i> -value	[0.021]	[0.399]	[0.028]	[0.061]	[0.061]	[0.062]	[0.844]
Adjusted <i>p</i> -value	<0.096>	<0.154>	<0.096>	<0.096>	<0.096>	<0.096>	<0.318>
Mean of control	10.4	2.7	6.9	210	1.20	3.695	24.77
Observations	182	179	175	174	191	18,489	6,560

Notes: Each column is a separate model of difference in means estimated by OLS. Estimated effects expressed as semi-elasticities for models (1)-(4) and (6). Standard errors in parentheses; one-sided *p*-values in brackets ($H_0 : \beta \leq 0$); adjusted *p*-values for multiple hypothesis testing (Romano & Wolf) in angled brackets. Mean of control group expressed in USD for models (1),(3),(6), and (7); number of clients in column (4).