Title

Can we improve quality of care in private health sectors? Evidence from a randomized field experiment in Kenya

Author

Kwan, Ada Ting Ting

Publication Date

2020

Can we improve quality of care in private health sectors?
Evidence from a randomized field experiment in Kenya

by

Ada Ting Ting Kwan

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 Jonathan T Kolstad
Professor Aprajit Mahajan

Summer 2020

Can we improve quality of care in private health sectors?
Evidence from a randomized field experiment in Kenya

Abstract

Can we improve quality of care in private health sectors?
Evidence from a randomized field experiment in Kenya

by

Ada Ting Ting Kwan

Doctor of Philosophy in Health Policy

University of California, Berkeley

Professor Paul J Gertler, Chair

At least half the world's 7.5 billion people lack access to essential health services, and many of the world's poor who seek care are at risk of catastrophic and impoverishing out-of-pocket spending. In low- and middle-income countries, well-established private sectors and low levels of quality, even in places with high access, contribute to the challenges of improving health care for the disadvantaged. To address these issues across Kenya, the 2012-2019 African Health Markets for Equity (AHME) program delivered a comprehensive package of interventions at private clinics to improve clinical decision making and to expand high quality health care access for the poor. By taking advantage of a randomized field experiment, we examine AHME's effects on quality of care and answer the question, *Can we improve quality of care in private health sectors?* We collected and analyzed unique data (including data from standardized patients, the state-of-the-art method to assess provider practice) to examine the program's effects on three quality of care dimensions: structures, processes, and health care outcomes. We find some significantly positive effects on structural quality, and surprisingly, AHME reduced correct care by 12% (p-value = 0.021). This relative reduction was consistent but not significant for rates of unnecessary testing and medicines, which dropped by 12.3% and 8.5% respectively, suggesting reduced waste. Since average knowledge of correct care was remarkably high (90% for diarrhea; 98% for malaria), effects may be because lowering quality is in the financial interest of these for-profit clinics, especially since patients and households did not recognize reductions in quality. We examine this hypothesis with a modified dictator game and find that the least altruistic providers at AHME clinics were the ones to reduce correct care, while charging 121% more than the least altruistic at non-AHME clinics. Across all clinics, we further uncover alarming deficits in laboratory quality and in the care given to poor clients compared to the non-poor. Since *highly competent* private providers give lower quality care to patients due to patient characteristics and provider preferences, the success of quality improvement in the private sector requires (1) stronger accountability with careful monitoring and (2) nuanced policies to account for patient and provider heterogeneity.

To Mam, Bàba, Popo & Yehyeh

Contents

Contents **ii**

List of Figures **iv**

List of Tables **v**

1 Introduction **1**

2 Background **6**
- 2.1 Study Setting . . . 6
- 2.2 The AHME Program . . . 7

3 Quality of Care Conceptual Framework **14**
- 3.1 Dimensions of Quality in the Conceptual Framework . . . 15
- 3.2 Research Questions . . . 16

4 Methods **17**
- 4.1 Experimental Design . . . 17
- 4.2 Program Compliance . . . 20
- 4.3 Quality of Care Measures and Data Sources . . . 21
- 4.4 Analytic Samples and Balance . . . 23
- 4.5 Identification Strategy . . . 29
- 4.6 Ethical Review . . . 29

5 Structural Quality **31**
- 5.1 Structural Quality Measurement . . . 31
- 5.2 Structural Quality Analysis . . . 34
- 5.3 Structural Quality Results . . . 36
- 5.4 Structural Quality Discussion and Conclusion . . . 37

6 Process Quality **38**
- 6.1 Process Quality Measurement and Analysis . . . 39
- 6.2 Process Quality Results . . . 46

6.3 Process Quality Discussion and Conclusion 54

7 Health Care Outcomes **57**

7.1 Measurement and Analysis of Health Care Outcomes 57

7.2 Health Care Outcome Results 59

7.3 Health Care Outcomes Discussion and Conclusion 60

8 Provider Preferences: Altruism vs. Self-interest **62**

8.1 Narratives from Fieldwork 62

8.2 Altruism vs. Self-Interest: Results from a Modified Dictator Game 64

9 Discussion **69**

9.1 Comparisons to Other Studies and Settings 69

9.2 Limitations of Our Quality Measures 70

9.3 Implications for Scale-Up, Program Design, and Research Design 71

10 Conclusion **72**

11 Research Team **73**

Bibliography **74**

Figures and Tables **80**

Appendices **132**

List of Figures

2.1. AHME Theory of Change 81
3.1. Quality of Care Conceptual Framework 82
4.1 Map: Kenya Population Density and AHME Evaluation Clinic Sample 83
4.2. AHME Evaluation Endline Surveys 83
5.1. Structural Quality Measurement: Criteria Used to Reduce Indicators 91
5.2. Conditional Density of Structural Quality Compliance by Dimension 96
5.3. AHME Effects on Structural Quality Compliance, by Indicator 98
5.4. Conditional Density of Capital Investments 100
6.1. Process Quality Measurement: AHME SP and Provider Survey Sampling 102
6.2. Effects of Gender on Process Quality: Asthma and Malaria SP Cases 112
6.3. Effects of Age on Process Quality: Asthma and Malaria SP Casess 113
6.4. Effects of Demanding Medicines on Process Quality: Childhood Diarrhea SP Case 114
6.5. Effects of Being Unmarried on Process Quality: Family Planning SP Case 115
6.6. Effects of Being Poor on Process Quality: Asthma SP Case 116
6.7. Effects of Being Poor on Process Quality: Malaria SP Case 117
6.8. Provider Knowledge and Practice for Correct Management and Diagnosis 119
8.1. Conditional Density of Social Preference Parameter: Alpha 128

List of Tables

4.1. Analytic Sample: Balance of Clinic Manager Characteristics at Endline 84
4.2. Analytic Sample: Balance of Baseline Clinic Characteristics at Endline 85
4.3. Analytic Sample: AHME and DHS Household Population Comparisons 86
4.4. Analytic Sample: Balance of Baseline Household Characteristics 87
4.5. Analytic Sample: Balance of Clinic Sample for Standardized Patient (SP) Data . 89
5.1. Structural Quality Measurement: Monitoring Data Differences across Implementers 90
5.2. Structural Quality Measurement: Monitoring Indicators Incorporated as Measures 92
5.3. Structural Quality Measurement: Compliance Measures 94
5.4. AHME Effects on Structural Quality Compliance by Dimension 95
5.5. AHME Effects on Capital Investments in Amenities and Equipment 99
6.1. Process Quality Measurement: Description of SP and Vignette Cases 101
6.2. Process Quality Measurement: Main Outcomes for SP and Vignette Cases 103
6.3. Process Quality Measurement: Family Planning Correct Management Components 103
6.4. Process Quality: Summary Statistics for SP Data 104
6.5. AHME Effects on Correct Management: SP Regression Models 105
6.6. AHME Effects on IMCI: Childhood Diarrhea SP Regression Models 105
6.7. AHME Effects on Any Unnecessary Lab Tests: SP Regression Models 106
6.8. AHME Effects on Any Unnecessary Medicines: SP Regression Models 106
6.9. Effects on Family Planning Treatment: SP Regression Models 107
6.10. Effects on Malaria Treatment: SP Regression Models 109
6.11. Effects on Childhood Diarrhea Treatment: SP Regression Models 111
6.12. Summary Statistics from Provider Survey Respondents 118
6.13. Effects on Knowledge of Correct Case Management 120
7.1. AHME Effects on Client Perceptions of Amenities: SP Models 121
7.2. AHME Effect on Client Perceptions of Amenities: Client Exit Interview Models . 122
7.3. AHME Effects on SP Satisfaction Rating (z-score) 123
7.4. AHME Effects on Providers Did a Good Job Explaining 123
7.5. AHME Effect on Household Perceptions 124
7.6. AHME Effects on Client Experience 125
8.1. Descriptive Statistics from Modified Dictator Game 126
8.2. AHME Effects on Client Budget Share Allocated in a Modified Dictator Game . 127
8.3. AHME Effects on Correct Case Management: Least Altruistic 129

8.4. Heterogeneity of AHME Effects on Correct Case Management: Least Altruistic . 130
8.5. Heterogeneity of AHME Effects on Prices: Least Altruistic 131

Acknowledgments

First, I would like to acknowledge the invaluable support of my dissertation committee. To Professors Paul Gertler, Stefano Bertozzi, Jonathan Kolstad, and Aprajit Mahajan: many, many thanks for your critical guidance, thoughtful discussion, organization of my disorganization, and steadfast confidence in me. To Professors Gertler and Bertozzi: what a treat it's been to be mentored by the two co-PIs of my first project in this field.

My special regards go to the individuals who played key roles in data collection. Asante sana to all the SPs, Rita Cuckovich, Afke Jager, and IPA-Kenya (especially Asman Suleiman, Allison Stone, Ginger Golub, Grace Makana, Jared Otuke, Eunice Kioko) for their excellent fieldwork management, survey intellect, trust, and hospitality; to Catherine Wanjiru, John Wanangwe, Jacob Odipo, Daniel Waithaka for their advice, commitment, and time; and especially to Andrew Muriithi, Pheliciah Mwachofi, Purity Kimuru, Rodgers Kegode, and Salome Omondi for being a dream team.

Much of this research is the confluence of conversations with many inspiring people. To the Qutub team, I thank Jishnu Das, Madhukar Pai, Benjamin Daniels, Veena Das, Ranen Das, Vaibhav Saria, Charu Nanda, Purshottam, Rajan Singh, and others at ISERDD for their ongoing lessons and conversations, agile mindsets, and all forms of support and generosities–related and unrelated–to methods and measures. For their intellectual inputs, discussions, and support throughout this work, my immense gratitude and appreciation go to: Guadalupe Bedoya, Jody Boffa, Claire Boone, Drew Cameron, Laura Chioda, Calvin Chui, David Contreras-Loya, Amrita Daftary, Maria Dieci, Amy Dolinger, David Evans, Kevin Feeney, Catherine Goodman, Joshua Gruber, Sean Higgins, Chenyu Hu, Reshma Hussam, Pamela Jakiela, Shachar Kariv, Jessica King, Alice Kwan, Mylène Lagarde, Jing Li, Jenny Liu, Alexa Magyari, Nancy Morrison, Camilla Nore, Zachary Olson, Nicole Perales, Timothy Powell-Jackson, Angela Salomon, and Misha Zatsman. I thank members of the UC Berkeley Health Policy Department and the UC Berkeley Health Economics Group, the full roster at the Bullpen, Nel Druce from the United Kingdom Department for International Development (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 for thoughts and feedback.

Lastly, I am extremely grateful to BMGF and DFID for funding this research and to the UC Berkeley Graduate Division, Professor Hector Rodriguez, Professor William Dow, and the UC Berkeley Department of Health Policy and Management for their financial support, resources, and mentorship, without which this research would not have been possible.

I have learned so much from all of you throughout this endeavor. In many different ways, this has been a group effort. However, all errors and mistakes are my own.

Chapter 1

Introduction

Achieving equitable access to high quality care in low- and middle- income countries (LMICs) is a challenge made up of at least three very different components: achieving access, improving care provided to clients, and ensuring equity. At least half the world's 7.5 billion people do not have access to essential health services (Evans, Whitehead, and Diderichsen 2001). Studies from India, China, and Kenya find that those who seek care are met with low and varied levels of quality of care, even in places with high access (Das, Hammer, and Leonard 2008; Das et al. 2012; Currie, Lin, and Zhang 2011; Sylvia et al. 2014; Mohanan et al. 2015; Das et al. 2015; Daniels et al. 2017; Kwan et al. 2018). For the poor, the consequences can be severe: many of the world's poor who *do* have access to care are at risk of catastrophic and impoverishing out-of-pocket spending (World Health Organization 2017). The World Bank and the World Health Organization report that every year, nearly 100 million people are pushed into extreme poverty (the equivalent of living on US$1.90 or less a day) due to paying out-of-pocket for health services (World Health Organization 2017). Nearly all of these 100 million people reside in LMICs, some of which have private health sectors that are rapidly growing and substantial sources of care. In these settings, interventions that improve access to high quality care for the poor must focus on private sector engagement, but the situation is far from simple.

What is clear is that there are existing needs to protect the most disadvantaged and ensure they have access to high quality care. Access appears to be differential across predictors of health disparities (Wang and Luo 2005; World Health Organization 2017) with private and public health care providers seeing different types of patients, but provider competence appears to be highly varied within and across LMICs (Das et al. 2018). It is unknown whether disparities result from the disadvantaged lacking access to competent providers once care is sought or whether competent providers give low quality care to the disadvantaged.[1] For quality improvement, the extent of the former and the extent of the latter require drastically different designs for health reform.

In order to construct effective quality improvement programs, what would help is doc-

[1] Understanding this requires complex data that frankly is difficult to collect.

umenting the problem more precisely. For example, if the disadvantaged - considering age, income, social class, education level, gender, and ethnicity - suffer worse health outcomes, but experience delays in seeking care, then patient access is an issue. However, once care is sought and quality of care is low, this can result in more sickness and lead to financial and non-financial costs to patients, health systems, and economies.

Low and varied quality of care associated with health disparities can arise from two possible channels, the extent to which either occurs is largely unknown in LMIC settings. The first channel is structural in nature: areas may lack in the supply of highly competent providers. In this sense, quality improvement programs ought to focus on incentives that may not be aligned for high performing providers to allocate or locate themselves in regions where disadvantaged populations reside. For example, areas with higher rates of poverty and diseases may also have lower skilled or poorly performing providers.

The second channel is behavioral, which could be a combination of provider agency problems and preferences (Arrow 1963; Becker 2010). Consider the scenario where providers are consistent across geography in their skills and knowledge; however, they may treat patients differently at the point of care. This could explain the "know-do" gap (i.e., the phenomenon in which providers perform few of the essential actions that would lead to correct diagnosis and treatment for the benefit of patients, despite demonstrated knowledge when interviewed (Das et al. 2012; Mohanan et al. 2015; Das et al. 2015; Satyanarayana et al. 2015; Sylvia et al. 2017; Das et al. 2018)), as well as the low and heterogeneous levels of quality of care from audit studies conducted with state-of-the-art methods in primary care settings in China, India, Kenya, and South Africa (Currie, Lin, and Zhang 2011; Sylvia et al. 2014; Das et al. 2015; Daniels et al. 2017; Kwan et al. 2018; Currie, Lin, and Meng 2014; Satyanarayana et al. 2016; Christian et al. 2018). If behavioral issues largely contribute to varied quality of care, quality improvement programs ought to focus on intervening on provider norms or actions that result from agency problems or preferences that can create or perpetuate economic or social disparities in health.

In Kenya, the African Health Markets for Equity (AHME) program was designed to improve clinical decision making and to improve the conditions related to accessing high quality of care for the poor. To achieve this, the AHME program made heavy investments in Kenya's private health sector between 2012 and 2019. In this study, we assess the effects of AHME's comprehensive package of interventions on quality of care. Based on a quality of care framework, we examine three aspects of quality of care: structures, processes, and health care outcomes. In so doing, we answer the question, *Can we improve quality of care in private health sectors?*

To answer our question, we take advantage of the randomized experimental design implemented to examine the effects of the AHME program.[2] From private primary care clinics sampled from 35 of Kenya's 47 counties, we collected and analyzed a unique set of data derived from clinic-level surveys, household and patient exit interviews, 322 provider

[2]The AHME program was funded along with a qualitative and a quantitative evaluation, and this study reports data mainly collected during endline for the quantitative impact evaluation.

surveys, and 1195 standardized patient (SP) visits. In the quality of care literature, SPs are typically healthy individuals recruited and trained to portray pre-designed health scenarios at sampled health facilities. The SP method is the state-of-the-art method, increasingly considered the gold standard, for assessing provider practice during a one-time interaction with a health care provider (Sylvia et al. 2014; Das et al. 2012; Das et al. 2015; Daniels et al. 2017; Das et al. 2016; Kwan et al. 2018; King et al. 2019; Kwan et al. 2019). Because providers are visited by the "same patient" with the "same illness", the typical confounders arising from differential patient and case-mix are better controlled for than in other data used to assess quality. By introducing experimental variation to SP cases, we are able to examine the effects of patient characteristics (age, gender, being poor, demanding non-efficacious drugs, and being unmarried) on quality. In doing so, we conduct the first study that utilizes the "gold-standard" SP method to causally identify effects of a wide scope of patient characteristics on care quality. We also conduct one of the first randomized field experiments that thoroughly examines structures, processes, and health care outcomes in the context of a program or intervention.

We summarize our findings in the following paragraphs. First, for structural and process quality, we find that the improvements on some structural quality measures due to the program did not translate into better process quality outcomes. For structural quality, the AHME program improved the protection of patients' rights and clinical record keeping. On average, structural quality was low in clinics that did not receive the AHME program.

Turning to our process quality results, we surprisingly find that AHME significantly reduced correct care for outpatient services by 12% (p-value = 0.021) compared to the AHME control group mean of 63.6%. We also find consistent reductions due to the program in unnecessary care suggesting reduced waste: the use of unnecessary lab tests dropped 12.3% (p-value = 0.423) compared to the AHME control group mean of 15.4%, and unnecessary medicines by 8.5% (p-value = 0.139) compared to the AHME control group mean of 58.6%. Although these results were not significant at the 10% level, they were consistent with the observed reduction in correct case management. We eliminate the hypotheses that these effects on correct and unnecessary care can be attributed to inadequate knowledge or stock levels for necessary supplies: average knowledge of how to correctly manage cases among the same providers were very high (90% for diarrhea; 98% for malaria), and in contrast to other studies, we find very low rates of stockouts.

For measures related to health care outcomes, we find very few program effects on client satisfaction and client perceptions of amenities—some of which are confirmed by client experience (e.g., wait time and time spent with the provider). We further do not find any evidence that households perceived any difference in quality between AHME treatment and control clinics, meaning households in clinic catchment areas did not notice the reductions of correct and unnecessary care provided by AHME clinics. The inability of patients to recognize the reductions in process quality suggests that the market failure could be due to medical information asymmetry between providers and clients. We conjecture that lower quality may be in the financial interest of these for-profit clinics, particularly if patients cannot recognize and negotiate for better care provided to them.

To test this hypothesis, we examine fieldwork narratives from the SPs and data from a modified dictator game designed to elicit social preferences. In particular, we explore whether there was heterogeneity in AHME's effects across providers based on altruistic preferences. We find evidence that supports our hypothesis. By sequentially identifying providers among the 50%, 25%, and 20% least altruistic, we find that the most profit-driven providers at AHME clinics were the ones to reduce correct care.[3] Adjusting for this in our model demonstrates that providers falling within the 80%, 75%, and even 50% most altruistic category do not contribute to the significant reduction in correct care due to the AHME program. Further, we find that the 20% most profit-driven providers at AHME clinics charged 121% higher prices compared to the most profit-driven providers at non-AHME clinics. We conclude that provider preferences, particularly self-interest, can have a deleterious effect in private sector engagement programs that focus on quality improvement.

Further, for all clinics—regardless of AHME participation—we find two particularly striking results. First, we find alarming deficits in laboratory quality for malaria services (outside the scope of the AHME program), which have potentially harmful consequences for individual and public health. Our team of SPs that portrayed malaria cases conducted 379 successful visits at clinics in our evaluation sample. Of the 311 visits that resulted in a malaria test, SPs in 86 (27.7%) visits received a false positive result.[4] Among 173 visits that resulted in a malaria microscopy test, 64 (37.0%) resulted in a false positive. Since all SPs portraying the malaria case were confirmed negative at the beginning and end of fieldwork (lasting 22 days), we consider these true false positives.

Our second striking finding was that poor clients received significantly lower rates of both correct and unnecessary actions compared to the non-poor. More broadly, providers in our sample treated patients differently, on average, based on certain characteristics: individuals between ages 22-35 were more likely to receive unnecessary tests with each additional year of age, but there was no age effect for medicines; females received more unnecessary medicines, but not more tests than males; those demanding a non-efficacious, but harmful drug got more correct care than those who did not demand anything as well as those who demanded a non-efficacious harmless drug; and we observed care quality differences only in ruling out pregnancy by marital status for a family planning case scenario.

Since highly competent providers give lower quality care to patients based on certain patient characteristics and provider preferences, our findings suggest that the success of private sector engagement and quality improvement requires (1) much stronger accountability and careful monitoring, and (2) more nuanced policies accounting for heterogeneity among patients and providers. Policies will have to counteract disparities and identify ways to ensure the disadvantaged can access care that is not only affordable but also does not buy them lower levels of quality with detrimental financial or health-related consequences. Qual-

[3]Following the literature on social preferences, identifying the "least altruistic" is equivalent to identifying the "most self-interested." Motivated by narratives from fieldwork, we interpret the self-interested to be the most profit-oriented in our context of the private health sector.

[4]Eighty six (22.7%) of all 379 visits made by malaria SPs resulted in a positive test result regardless of having a test ordered.

ity improvement programs may also want to invest in identifying and/or changing norms, such as towards fair-mindedness or altruistic preferences.

This dissertation is organized as follows. We begin with background on the AHME program in Chapter 2. In Chapter 3, we describe the conceptual framework that guides our research questions on the effects of AHME on different types of quality of care: specifically, health care structures, processes, and outcomes. Chapter 4 details the experimental design of the AHME impact evaluation as it pertains to this study, AHME program compliance, methods, and ethical review. Corresponding to our quality of care framework, we then designate a chapter for each set of our research questions mapped to each type of quality of care. With each chapter containing its own section on methods (measurement, outcomes, and empirical approach), as well as results and discussion sections, Chapters 5, 6, and 7 focus on the effects of AHME on structural quality, process quality, and health care outcomes, respectively. Chapter 8 focuses on provider altruism and heterogeneous effects of the AHME program. Chapter 9 provides a discussion of our findings, including comparisons to other studies, our study limitations, and implications for policy. Chapter 10 closes, and Chapter 11 details the research team.

This research makes three main contributions to the literature. First, this study contributes to our understanding of the complexities involved in improving quality at a time when governments 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, to our knowledge - explore quality improvement and the roles of patient characteristics and perceptions, as well as provider competence, practice, and preferences. Second, this research contributes to the growing literature on the use of experimental audit studies to understand the extent agents treat consumers differently based on certain signals, such as age, gender, wealth, marital status (Bertrand, Chugh, and Mullainathan 2005; Bertrand and Duflo 2017; Arrow 1998; Stiglitz 1973). Third, this study methodologically contributes to the literature on quality of health care measurement. Most research studies on health services exclude multiple quality of care measures and/or suffer from endogeneity problems when estimating quality. This is often because of measurement and data issues, as well as budget constraints. Because this study uses a randomized field experiment of an at-scale program and introduces exogenous variation in patient characteristics with the SP method, 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 (Gertler and Waldman 1992; Hurley 2000; Fitzpatrick and Tumlinson 2017).

Chapter 2

Background

2.1 Study Setting

This study focuses on primary care clinics in Kenya's private health sector. Kenya is a lower-middle-income country located in East Africa with a population of 49.7 million (2017) and US$1,507.81 GDP per capita.[1] There are three main reasons why Kenya is an ideal study setting to understand the impact of health interventions and private sector engagement on quality of care.

First, for health care, there is a growing need to understand access to quality care across Kenya, particularly for the disadvantaged. In addition to the current administration's commitment towards access to quality health care as a constitutional right, the Government of Kenya has begun engaging the private sector in Universal Health Coverage (UHC) and National Hospital Insurance Fund (NHIF) activities, including a social health insurance scheme aimed to protect the poor. Thus, understanding quality of care and how it relates to underserved and disadvantaged individuals is relevant for national policy as well as international health policy (World Health Organization 2017; Kruk et al. 2018).

Published in 2017, a quality of care study using standardized patients (SPs) found variations in quality of care across public and private facilities and across health conditions in Kenya's capital city of Nairobi (Daniels et al. 2017). Nairobi has the highest living standards in the country, and it is unknown the extent of provider practice variation across the country, where rates of poverty are higher. Our study focuses on clinics across Kenya's private health sector, which contains 40% (4370) of Kenya's 10,994 health facilities.[2]

Second, Kenya has documented inequalities and health disparities that vary across geography. In 2016, 36.1% of the population was living below the national poverty line, [3] and the majority reside in rural areas. In 2006, 2.5 million individuals who were poor lived in urban areas, and 14.1 million lived in rural areas. A poverty and inequality assessment

[1]World Bank, wdi.worldbank.org/

[2]Kenya Master Health Facility List, http://kmhfl.health.go.ke/

[3]World Bank Group, https://data.worldbank.org/country/kenya

report by the World Bank finds significant differences across regions and counties for poverty and attributes the spatial variation to differences in climate, agro-ecology, and different levels of access to services(World Bank 2008). Many of elements of these predictors have persisted from British colonial rule and settlement.[4] The report has several other findings. First, nearly one out of 5 households would not meet the cost of minimal food bundles if they spent their entire budget on food. Second, although income growth did not occur as rapidly for the poor as the wealthiest, access to key services improved more rapidly for the poor, reducing existing gaps in access.

Third, the African Health Markets for Equity (AHME) program was implemented at-scale across Kenya for over six years. This program engaged the private health sector by delivering services to households within clinic catchment areas. To assess the impact of this program, an impact evaluation was conducted between 2013 and 2019 with baseline and endline data collection activities. With the advantage of the randomized experimental design of the program, the data from the impact evaluation offers a unique opportunity to causally assess whether we can improve quality of care.

2.2 The AHME Program

The AHME program was a comprehensive package of interventions designed to improve access to high quality primary health care for the poor delivered by financially sustainable private clinics. Implemented over more than six years from November 2012 through March 2019 in Kenya, the program had three main goals: (1) to increase access to and utilization of health care by the poor, (2) to expand coverage of priority health technologies and interventions for the poor, and (3) to improve the quality of care provided by private health care providers (clinics) and their business viability. The program jointly addressed the supply side, the demand side, and policy through a coordinated package of interventions, working through networks of franchised health care clinics. Clinics franchised through the program delivered specified health services under a common brand, with a promise of quality assurance.

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

[4]Kenya gained independence in 1963.

The AHME Program's Theory of Change

The design of the AHME program rested on the belief that large, interdependent systems (supply, demand, policy) need to be addressed in tandem to create a health market that functions well for the poor (figure 2.1). The AHME theory of change proposed five necessary conditions for health markets to work well for the poor: (1) the poor must be enrolled in national health insurance, (2) primary health care must be covered by national health insurance, (3) clinics must accept national health insurance, (4) high quality services must be available through those clinics, and (5) clinics must have viable financial business models. If any of these five conditions is not addressed, health markets will not be able to adequately function for the poor. The program aimed to jointly address supply side, demand side, and policy factors through corresponding interventions:

- ***Supply-Side Interventions:*** To improve the supply of high quality health care services, AHME invested in scaling up networks of franchised private clinics. AHME aimed to expand the existing scope of services, to improve quality of care, and to improve business financial sustainability through access to capital, better management, and accreditation with national health insurance systems.
- ***Demand-Side Interventions:*** To increase the demand for health care services, AHME aimed to remove barriers for clinics to enroll in NHIF, which gave insured patients the ability to finance their own care through NHIF. AHME also helped franchised clinics develop and institute marketing and client outreach plans.
- ***Policy Interventions:*** The health policy component advocated for pro-poor enrollment policies for insurance, as well as cost-effective packages of health care services that would be affordable and beneficial for the poor.

Interventions Included in the AHME Package

AHME core activities built on social franchising carried out by MSK under the Amua brand and 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. This section summarizes the important aspects of the interventions in the AHME package.

Social Franchising

Social franchising binds together a large number of small, independent providers (clinics) in order to leverage scale to improve supply chains, exploit joint brand advertising, and strengthen worker training and supervision. Social franchises are able to scale rapidly, decrease transaction costs, standardize services, collectively negotiate financial reimbursement mechanisms, and replicate best practices among a large group. Brand advertising and education programs help promote franchise products and strengthen critical links to low-income consumers.

In Kenya, franchising for the AHME program was overseen by two social franchise networks: MSK's Amua network and PSK's Tunza network. In order to be franchised, clinics were required to meet minimum standards, including proper licensing, no memberships in other networks, regular and adequate operating hours, as well as other requirements for services offered. The specific services offered as part of franchising included the following:

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

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

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

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

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

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

SafeCare

SafeCare is a certification program that focuses on improving the structural 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. 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). Accredited by the International Society for Quality in Health Care (ISQua), SafeCare's primary goal to "increase trust between patients, providers, financers, and governments by making medical and financial risks more transparent and identifying quality gaps, paving the way for sustainable investment in quality, required to improve scale and scope of healthcare delivery in lower and middle income countries" (PharmAccess Group 2017). During the AHME program, clinics were first assessed across various service areas (PharmAccess 2017a, 2017b) from which MSK and PSK developed customized Quality Improvement Plans (QIP) for each clinic (PharmAccess 2017c, 2017d). Subsequently, MSK and PSK conducted routine visits at clinics to check on the progress of QIP implementation. Once the implementing partners confirmed significant progress had been made on the QIP, a reassessment was carried out and the QIP was revised. The SafeCafe program focused on 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 Healthcare 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.[5] 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.

Anecdotally, clinics participating in SafeCare in Kenya were also motivated to participate as a means to improve readiness for NHIF accreditation standards. "[Providers] felt that SafeCare made it easier for them to both become NHIF-accredited (most applicable to providers in Kenya) and to renew their accreditation...; in this regard, providers often suggested that SafeCare helped them 'put things in place' for the accreditation process" (Suchman, Montagu, and Seefeld 2020, p.16).

Business Support

The main aim of Business Support in AHME was to enable clinics to be run as profitable professional businesses. The franchisers first assessed clinic business practices using a Business Assessment tool (Population Services Kenya 2016), from which they helped each clinic develop a Business Improvement Plan (BIP) (AHME Program 2016). MSK and PSK organized multi-day business training workshops and helped to set up business structures and systems, such as keeping financial records, doing stock management and developing strategic plans. The franchisers also gave guidance on business development and facilitated financing for expansion by helping obtaining bank loans.

The Business Assessment tool measures the quality of clinic business structures and systems through five categories, each containing 3 to 7 criteria. Each criterion is ranked Severe, More Severe, or Most Severe as a way to determine which item should be prioritized as necessary to improve clinic operations. The categories are:

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

[5] 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.

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.

A Business QIP 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.

NHIF Empanelment Support

As part of the business support component, AHME helped clinics become accredited by NHIF in order to be able to treat insured clients and accept NHIF insurance payments for those services. The hypothesis was that increasing the number of clinics accepting NHIF insurance would increase the potential client base of clinics and improve the affordability of care. The NHIF Empanelment Support program was designed to educate clinic owners about NHIF empanelment, and assist clinics interested in empanelment throughout the application process.

Prior to AHME, while clinic interest in empanelment was high and the number of clients using NHIF insurance in Kenya was increasing, the numbers of private clinics in Kenya accepting NHIF insurance was low. Only 12% of the 123 AHME treatment group clinics were accepting NHIF insurance prior to AHME. This was likely to be partially due to the bureaucratic, time-consuming process to become empaneled.

Chapter 3

Quality of Care Conceptual Framework

In this chapter, we describe a conceptual framework for quality of care based on quality of care literature. We use this framework to separate types of quality measures and to construct research questions on the impact of the AHME program on quality of care. At the end of this chapter, we pose research questions to understand the impact of the AHME program on quality of care.

The AHME program sought to ensure that health care markets function for the poor by improving the supply of health care, the demand for health care, and health policy in the private sector. Central to its supply-side strategy was a package of interventions designed to improve quality of care in private health clinics. Access to high quality care is necessary to translate solvent business practices and insurance coverage into better health outcomes for the poor.

In order to improve quality of care at private clinics in Kenya, the AHME program implemented four clinic-level interventions: (1) Social Franchising, (2) SafeCare quality improvement, (3) Business Support, and (4) National Hospital Insurance Fund (NHIF) Empanelment Support, as discussed in Chapter 2. Social Franchising works to ensure standardized quality of care delivered under a common brand. To standardize across franchises, the two franchisers — Marie Stopes Kenya (MSK) and Population Services Kenya (PSK) — provided clinics with clinical guidelines, training, supervision, supply chain management, and other supportive services to meet minimum standards, which were assessed regularly via internal quality audits. Through the Business Support intervention, MSK and PSK helped clinics develop Business Improvement Plans that included investments in clinic, stock management, safety procedures, and clinic cleanliness. SafeCare provided customized and iterative Quality Improvement Plans to clinics based on regular assessments. The results of those assessments qualified clinics for a tiered certification program emblematic of achieving consecutive levels of quality of care standards. This certification program strongly resembled NHIF accreditation standards. Through NHIF Empanelment Support, MSK and PSK supported clinics in meeting NHIF-required quality standards by conducting preassessments and self-assessments using the NHIF checklist and developing quality improvement plans focused on deficiencies identified through the NHIF checklist.

Quality of care is difficult to measure and assess (Donabedian 1978, 1985, 1988). Different definitions and dimensions exist in clinical guidelines as well as the quality of care literature. For example, quality of care or a dimension of quality of care can be considered effective (good, as in timely and accurate) or non-effective (which can be either harmless or harmful). To understand the impact of AHME on quality of care, we draw from Donabedian's conceptual framework for quality of care (see figure 3.1) and construct metrics based on national health guidelines (Donabedian 1966, 1978, 1988). The framework categorizes dimensions of quality of care into three measurements: *health care structures*, *processes*, and *outcomes*. Health care structures, processes, and outcomes play a role in the interaction between an individual's health and multiple actors: the client, the provider (health care practitioners, including doctors, nurses, clinical officers, and midwives), and the clinic (including the environment and resources). The framework further separates each type of quality measurement into *interpersonal* and *technical components*.

3.1 Dimensions of Quality in the Conceptual Framework

In this section, based on the Donabedian Conceptual Framework, we describe the three main quality of care measurements (health care structures, processes, and outcomes) and the two components of quality for each measurement (interpersonal and technical).

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

Second, ***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 based on the client's symptoms and responses to history questions, and follow up by prescribing efficacious medicine for the client's ailment. (Differential diagnosis is the process of differentiating between two or more health conditions that share similar history or symptoms.) 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. Whether a client demands correct or incorrect services (versus lacking correct knowledge or versus full compliance) or is older (versus younger), female (versus male), unmarried (versus married), or poor (versus wealthy) may play a role in a provider's diagnostic and treatment decision making. Similarly, provider characteristics may influence process quality outcomes.

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

For each of the three dimensions of quality of care measurements, the framework further separates 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 and technical components of quality, respectively.

3.2 Research Questions

Based on the conceptual framework, our research questions to assess the effects of AHME on quality of care are organized by the three main quality measurements: health care structures, processes, and outcomes. The three research questions about structural quality are:

1. How can a structured survey measure structural quality?
2. What was the effect of the AHME program on structural quality?
3. What was the effect of the AHME program on capital investments in physical assets?

The three research questions about process quality are:

1. To what extent did the AHME program have an impact on correct actions related to process quality?
2. What was the effect of the AHME program on nonessential or unnecessary actions related to process quality?
3. What was the effect of the AHME program for different services and clients with different characteristics?

The three research questions about health care outcomes are:

1. What was the effect of the AHME program on client perceptions of amenities and client satisfaction?
2. What was the effect of AHME on the perception of clinic quality among households?

Chapter 4

Methods

Data used in this study are from quality of care surveys embedded into the AHME impact evaluation. The quasi-randomized AHME impact evaluation was conducted across Kenya between 2013-2019 and consists of two main waves of data collection (conducted at baseline and at endline) and tracking data across that time period. This experimental design allows us to conduct intent-to-treat (ITT) analyses to understand AHME's effects on quality of care. For this study, we primarily analyze endline patient-, provider, clinic, and household-level data we collected between 2018-2019 across n=232 primary care clinics that comprise the AHME evaluation clinic sample. The AHME evaluation clinic sample (n=232) is distributed across 35 of Kenya's 47 counties. Figure 4.1 depicts maps of the AHME evaluation clnic sample and Kenya population density. Figure 4.2 shows the AHME evaluation endline surveys mapped to various measures of interest.

In this chapter, we detail our methods to answer our research questions. The chapter first describes the experimental design of the AHME impact evaluation and the AHME program compliance. Then, we detail the data sources used to assess quality of care; the analytic samples, including details on balance achieved at baseline; our identification strategy, and ethical clearance and considerations related to this study. Details on the larger AHME impact evaluation are included as they pertain to this study, and the reader can access Gertler et al. (2020) for additional details.

Corresponding to our quality of care framework, Chapters 5, 6, and 7 focus on the effects of AHME on structural quality, process quality, and health care outcomes, respectively. Because of the distinct methods required to assess each type of quality of care, each of these chapters contains its own section on methods (measurement, outcomes, and empirical approach), as well as results and discussion sections.

4.1 Experimental Design

This section describes the experimental design and clinic selection that guides our ITT analysis. Before the AHME program began in 2012, we mapped all clinics in the evaluation study areas (35 of Kenya's 47 counties) with the goal of randomizing clinics eligible for the program into treatment and control groups. Clinics were first identified for mapping using four sources of information: (1) official government list of private clinics in the country, (2) clinics belonging to a major professional health associate (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. Government clinics (public clinics and hospitals), faith-based clinics (identified by clinic name), and clinics that were identified as franchised (by franchise branding on clinic exterior) were removed from the sampling frame generated from the aforementioned information sources.

A pre-screening and baseline survey among remaining clinics was administered to remaining clinics. The purpose of this was to exclude clinics that the implementing partners indicated were not eligible for franchise services or the AHME set of interventions. Clinics that met the basic eligibility criteria still varied in their "level of eligibility" based on their existing capacity and suitability for franchising services and AHME interventions. Using additional criteria, created in a collaborative manner with MSK and PSK, clinics were further categorized into groups based on how likely they were to be eligible ("eligibility tiers") using data collected through the baseline survey instrument. After clinics were categorized by eligibility tier, the research team conducted a stratified randomization of eligible clinics. For the randomization process, clinics were grouped into their eligibility tiers within a county based on partner-provided criteria and data from the baseline survey, randomly ordered within strata (groups within which randomization would occur), and then randomly assigned to treatment (eligible to be offered AHME franchising immediately) or control (not eligible to be offered AHME until the completion of the study).

After randomization procedures were completed, we provided MSK and PSK with partner-specific recruitment lists indicating the order in which clinics on their lists were to be approached for screening ("sensitization") and recruitment. Once randomization procedures had been completed, MSK and PSK began engaging clinics on their lists and proceeded with their respective recruitment procedures in October 2013. The second round of screening and recruitment by the partners served to identify clinics that were eligible for franchising and AHME interventions in the treatment arm. Eligible clinics that were invited to join either franchise ("ever franchised") were considered part of the evaluation sample in the treatment arm. Consistent with the ITT assumptions applied in this study, these ever-franchised clinics were considered part of the final evaluation sample regardless of whether they completed the franchise enrollment process or maintained their franchise enrollment status for the entirety of the study period (for any reason).

In the remainder of this section, we provide further details on constructing the clinic evaluation sample, clinic mapping and screening procedures, treatment arm recruitment, and control honing. If curious, the reader may access Gertler et al. (2020) for further details on the experimental design, including recruitment eligibility and randomization.

Clinic Evaluation Sample

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

Clinic Mapping and Screening

During the mapping process, more than 7000 potential clinics were identified nationally from government lists, partner recommendations, and field team observations. From this universe of clinics, 4216 were excluded for various reasons, including already franchised, faith-based or public clinic, and proximity to an existing franchised clinic. The remaining 3467 clinics were screened by the evaluation team using standardized surveys (pre-screening and baseline clinic surveys). Of these, 472 clinics were excluded for not meeting initial screening criteria. The remaining 3002 clinics were randomized into treatment and control clinics, assigned to MSK or PSK, and randomly ordered using the stratified design. Clinics considered ineligible were removed from the sample using the original criteria prior to the first recruitment round (N = 836; treatment: N = 418; control: N = 418) and the updated criteria prior to the second round of recruitment (N= 486; treatment: N=258; control: N = 228). The remaining 1680 randomly ordered and randomly assigned clinics were considered eligible (treatment arm: N = 851; control arm: N = 829). Recruitment lists were prepared for MSK and PSK to visit the 851 clinics in the treatment arm.

Treatment Arm Recruitment

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

Control Honing

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

4.2 Program Compliance

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

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

The results of this monitoring demonstrate that implementation of the AHME package of interventions was done well at the clinic level. Approximately a quarter of clinics disenrolled from AHME over time. Among the treatment clinics who remained until the end of the program, 97% of those that were ever franchised were visited at least once a quarter by a Social Franchising Coordinator, and 74% were visited monthly or more often. At least 90% of clinics in the treatment group had exposure to Social Franchising, for more than 18 months, while 80% had SafeCare and Business Support for the same period. Approximately 76% were offered help with NHIF empanelment. An additional 15% of treatment clinics were already empaneled at the beginning of the study, and at the end of the AHME program, 41% of treatment group clinics were empaneled in NHIF. An additional 43% were taking part in the empanelment process but were not yet empaneled.

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

4.3 Quality of Care Measures and Data Sources

To answer our question on whether we can improve quality of care, 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. This section aims to equip the reader with a brief description of quality of care methods in order for her or him to better interpret the nuances and limitations of measuring quality of care dimensions in this study. To summarize, popular methodologies to understand quality of care include: patient exit interviews, provider vignettes, direct observation, medical record extraction, and standardized patients (SPs). After considering the various methods to collect quality of care data, this study will use a combination of interview and direct observation in clinic surveys to assess structural quality, SPs to assess provider practice among individuals portraying themselves as patients, exit interviews to assess provider practice among real patients, provider surveys to capture provider characteristics, provider vignettes to assess provider knowledge, and a modified dictator game to capture provider preferences.

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

The purpose of our quality of care surveys is to capture aspects that define health care structures, processes, and outcomes from different perspectives. To do this, we implemented the following five surveys at endline: (1) a clinic survey; (2) a standardized patients survey; (3) a provider survey; (4) client exit interviews; and (5) household surveys. The surveys are briefly described below and in respective chapters 5, 6, and 7. More details on the surveys can be found in Gertler et al. (2020), including the mapping of AHME theory of change to quality of care outcomes and indicators and details on the purpose of the AHME impact evaluation's surveys, recruitment and indicators.

Clinic Survey

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

Standardized Patients

To measure process quality, we collected and analyzed primary data from N=1195 standardized patient (SP). 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. We designed the SP surveys to capture (1) the remaining four monitoring indicators (all four were identified by IRT and were process quality measures for the management of childhood illnesses); and (2) health care process and outcome measures related to the main outpatient services (child curative, adult curative, and family planning services) covered by the AHME intervention. In order to identify the causal effect of various client characteristics on quality of care, we also conducted experiments in the SP surveys that varied one characteristic in SP case presentation for each case scenario and randomly assigned them to the AHME evaluation clinic sample. 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 Chapter 6.

Provider Survey

To complement our main analyses on structural and process quality, 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 written case scenarios matching three SP case scenarios. The vignettes are designed to capture provider knowledge of case management and are administered by an enumerator in an interactive interview. For example, in one case scenario, if the respondent states he would take the client's temperature, the enumerator responds with the pre-scripted temperature measurement, but unlike a client, the enumerator is known by the provider to be a data collector, guides the provider through the interview, and also notes down responses. We also designed a real-stakes, adapted dictator game to capture the respondent's social preferences (altruism).

Exit and Household Interviews

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

4.4 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 describe the process of recruiting clinics to the study from a universe of private Kenyan health care clinics and explain the creation of the final clinic evaluation and analytic samples. We provide evidence supporting the effectiveness of randomization at creating clinic treatment arms that are, on average, exchangeable (balanced with respect to observable baseline characteristics). We provide similar results for the recruitment and sampling of the households and quality of care analytic populations from the 232 clinics in the AHME impact evaluation's clinic sample.

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

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. The following subsections describe survey analytic samples and tests for balance at baseline.

Clinic Analytic Sample

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

Baseline Balance of and Description of the Clinic Sample

We compared key baseline characteristics across treatment and control arms for the clinic sample, including manager characteristics, services offered, and financial and operational indicators. Table 4.1 reports treatment arm specific means, difference in means, and associated t-test results for select baseline clinic characteristics. Overall, we find that on average, clinic owners and managers are comparable across treatment arms. A large proportion of clinic owners also operate as managers (treatment: 67%; control 62%). Managers are comparable in terms of their qualifications (87% of managers in treatment clinics; 89% of managers in control clinics had a clinical degree); demographic characteristics (gender, age); and experience (tenure, positions at other clinics) (see table 4.1).

Table 4.2 reports means and standard errors for the types of services provided at the clinics, and other key financial and operational characteristics by treatment arm. Of the 199 clinics in the analytic sample, we find that clinics are similar on the following characteristics: utilization levels, number of years the clinic has been operating under the same management (ownership); and the proportion of clinics that provided services such as antenatal care, labor and delivery, postnatal care, tuberculosis testing and treatment, and hospitalization (inpatient services). Clinics in the sample also had similar unit costs, profit margins, and comparable levels of contractual arrangements with the government (National Hospital Insurance Fund, NHIF) and private companies. The only two differences that were found are related to the scope of child health services offered between treatment and control groups: the proportion of clinics that offered child immunization services (30% in the treatment group vs. 46% in the control group, p-value<0.05), and the proportion of clinics that offered well-baby check-ups (62% in the treatment group vs. 73% in the control group, p-value<0.10).

Household Sample

Household Evaluation Sample

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

Household Sample Recruitment from the Clinic Evaluation Sample

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

Description of AHME Household Sample

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

Comparison of AHME and DHS Households in 2015/2016

Compared to DHS households, AHME households in the same province, that accessed care at AHME clinics in 2015 or 2016, on average, were slightly more likely to be urban (AHME: 42%; DHS: 41%) and to have more household members (median household members, AHME: 5; DHS: 3) (table 4.3). Heads of households were more likely to be male in the AHME sample (AHME: 89%; DHS 64%), younger (median years of age, AHME: 35; DHS 39), and more likely to have any education (AHME: 96%; DHS 87%). In 2015 and 2016, households in the AHME sample were also more likely to have a finished floor (AHME: 62%; DHS 58%) and to report having access to an improved source of water (AHME: 78%; DHS 70%) and their own, non-shared, improved sanitation facility (AHME: 47%; DHS 28%) (table 4.3). The distribution of household asset wealth among the AHME and DHS populations based on asset ownership in 2015 and 2016 are similar; however, the AHME distribution is truncated at both tails, suggesting it has proportionally fewer "very poor" and "very rich" households compared to representative households in the same provinces. The increased peak of the AHME distribution also suggests a higher concentration of households that had median and above-average wealth at baseline, compared to DHS households in the same regions. We further found that the DHS households in the regions where AHME was rolled out followed the same quintile distribution as the national population (20% in each quintile). Households included in the AHME endline sample were less likely to be in the first (poorest) quintile (AHME: 13%; DHS 20%) and fifth (wealthiest) quintile (AHME: 10%; DHS 20%), and were more likely to have median levels of wealth, with an increasing trend toward the wealthier quintiles (quintile 2: 24%; quintile 3: 25%; quintile 4: 26%).

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

Baseline Balance of Household Sample

Within the AHME sample, we evaluated the balance of demographic, socioeconomic, and infrastructure indicators collected at baseline. Table 4.4 reports the means (proportions) and standard errors for treatment and control households. Differences in means between the two groups are evaluated using t-tests for single indicators, and F-statistics test the joint statistical significance for each group of outcomes (e.g., demographics, infrastructure, socioeconomic indicators). We find that the endline household evaluation sample overall is well balanced on observable baseline characteristics, suggesting that randomization was effective. This provides evidence that the household control arm is a good counterfactual for the AHME treatment group and supports use of the ITT parameter throughout to provide unbiased estimates of the impact of treatment on household outcomes.

Quality of Care Sample

Quality of Care Sample Recruitment

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

Description of Quality of Care Sample

Analyses using the endline clinic surveys, patient exit interviews, and the provider survey rely on the full clinic evaluation survey sample. The household data analyses related to quality of care rely on the full household sample at endline. Analyses using SP data relied on clinics from the full evaluation sample that were both eligible to receive clients for our study's SP case scenarios and had consented to previous AHME evaluation endline surveys. This section summarizes the various quality of care samples; however, for more details on the different samples, design, and data, the reader ought to refer to respective Chapters 5, 6, and 7.

We collected quality of care measures from the endline clinic survey and patient exit interviews from the clinic sample (N = 199), which is described in section 5.1. The endline household survey relied on the endline household survey sample. As for the SP and provider surveys, figure 6.1 depicts the quality of care analytic samples by AHME treatment assignment for the SP and provider surveys (described further in Chapter 6). For the SP sample, we restricted the 232 randomized clinics from the full evaluation sample by excluding clinics that did not consent to previous surveys (N=3), were located in a conflict region (N=1), and were ineligible for SP cases (N=4). In total, 8 of the 232 clinics were excluded or ineligible for SP surveys. Between February and May 2020, after data collection for endline exit interviews was completed, SPs attempted visits at the remaining 224 clinics (treatment clinics: N=120; control clinics: N=104). All of these visits were conducted by SPs trained to portray pre-scripted cases seeking outpatient walk-in visits for family planning, child curative, and adult curative services.

During the SP data collection, supervisors and SPs worked to identify providers seen at the clinics. We constructed a provider survey sampling frame from the list of identified providers, and the data collection team was instructed to interview the full list. The group of providers that were not successfully identified were classified as "unidentified" providers. Among the treatment group, 651 SP visits seen by 221 identified and 128 unidentified providers were successfully conducted at 114 treatment clinics. Among the control group, 544 SP visits seen by 179 identified and 101 unidentified providers were successfully conducted at 97 clinics. We analyzed all 1195 successful SP visits from 211 clinics.[1]

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 (see table 4.5). These clinic indicators include client utilization, ownership, services provided, average costs and average revenue per patient, profit margins, whether the clinic accepted NHIF, whether the business paid for staff, and whether the clinic sold medicines to the public. For clinic indicators collected at baseline, we see that the N=211 clinics in the SP sample were balanced on 15 of the 17 variables at the 5% significance level. By chance, some characteristics were statistically different between the AHME treatment and control households, at the 5% significance level. Notably, whether the clinic provides child immunization (control proportion = 0.48, SE 0.05 vs. treatment proportion = 0.33, SE 0.04) and whether the clinic provides well-baby check-ups (control

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

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.

4.5 Identification Strategy

In this study, we use data from SPs to address typical endogeneity issues with patient data, such as patient sorting and case-mix issues. This method is considered the gold standard for measuring provider practice in a one-time interaction. We introduce exogenous variation to standard case scenarios presented by SPs to causally identify effects on specific patient characteristics. Since SP data does not account for different case-mixes, SP data is complemented by real patient exit interviews conducted at the same private clinics. However, because we are examining primary care services (childhood illnesses, family planning adult respiratory health, adult malaria) in a sample of primary care clinics, we did not expect and did not find provider specialties.

Whether a provider provides high quality care (i.e., performs timely and accurate clinical actions) during an interaction with a patient is due to many factors, including but not limited to competence, training, stock of medicines, and preferences. In this sense, understanding training and structural constraints is important. To do this, we analyze structural quality, as well as provider knowledge, provider preferences, and other characteristics collected from vignettes and provider surveys. These surveys were administered among clinic owners and practicing providers at the same clinics as the exit interviews and SP visits. Second, additional information on whether a clinic has stockouts and other characteristics are collected from SP visits and clinic-level surveys.

4.6 Ethical Review

This study was granted clearance by the ethics committees within the AHME quantitative evaluation at Kenya Medical Research Institute (No. KEMRI/RES/7/3/1; NON-SSC PROTOCOL NO. 372) and the Human Subjects Committee for Innovations for Poverty Action IRB-USA (IPA IRB Protocol 1085). All the SPs in this study were hired as field staff and participated in a 3-week training, 2-week pilot and are 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. Detailed information can be found in the appendix 6.A1.

Similar to other SP studies with similar designs and embedded in an intervention (Kwan et al. 2018; Kwan et al. 2019), we sought a waiver of provider informed consent to conduct the SP study. The request for a waiver was based on a recent study commissioned by the United States Department of Health and Human Services to assess the ethics of simulated patient studies (Rhodes and Miller 2012). Supported by a pilot study conducted in Nairobi which validated the SP method in the Kenyan context (Daniels et al. 2017), both ethics committees approved the waiver request within the AHME evaluation study (the approved waiver request is included in appendices 6.A1.2 and 6.A1.3), since (1) combining informed consent with the congregation of providers during trainings and the implementation of interventions during the study period posed threats to the scientific validity of the study objectives as well as to the risk of SP detection, and (2) there is no more than minimal risk of participation to the SPs or providers, as reported in the Nairobi SP pilot and validation study (Daniels et al. 2017). The appendix includes an exit questionnaire and case scenario for childhood diarrhea; however, all other questionnaires and case scripts are available upon request.

Chapter 5

Structural Quality

This chapter focuses on the second type of quality of care from our framework: structural quality. Specifically, we seek to answer the following three research questions related to structural quality:

1. How can a structured survey measure structural quality?
2. What was the effect of the AHME program on structural quality?
3. What was the effect of the AHME program on capital investments in physical assets?

We describe how we leveraged internal monitoring data collected by MSK, PSK, and SafeCare to measure technical and interpersonal quality using structured surveys at endline. We reduced the pool of variables from 3,178 internal monitoring indicators to 32, and incorporated those measures into our clinic surveys to collect at endline. We constructed two indices to measure structures to facilitate interpersonal care and structures to facilitate technical care, and estimated the effect of AHME program assignment on structural quality. We find some significantly positive effects of the AHME program on interpersonal and technical components of structural quality. We found particularly large improvements in protecting patients' rights and keeping clinical records. The AHME program significantly increased the likelihood that a clinic would offer informed consent, display a patients' rights chart, keep staff files with credentials, and keep a list of referral sites. The remainder of the chapter details structural quality measurement and analysis, results, and discussion.

5.1 Structural Quality Measurement

We leveraged internal monitoring data collected by MSK, PSK, and SafeCare to measure technical and interpersonal quality using structured surveys at endline. Internal monitoring includes (1) internal quality audits by MSK and PSK, as well as (2) service area assessments by SafeCare (see chapter 2 for further details).

In 2018, implementing partners shared administrative records of clinic assessments with the AHME evaluation team. In each assessment, an auditor from the implementing organization evaluated clinics in the program on a set of indicators. The relevant set for a clinic depended on the implementing partner, services provided at the clinic, and status of advancement in the franchising program. These AHME clinics were each assessed one to four times from May 2012 to September 2018. Table 5.1 summarizes key differences in the data shared by implementers. PSK collected information on the most indicators (1,591 of 3,178 total variables) and evaluated each indicator on a binary scale (yes/pass or no/fail). In contrast, MSK and SafeCare used ordinal scales. SafeCare scored each indicator as not, partially, or fully compliant, while MSK scored each as not in place, partially achieved, or fully achieved. Both the monitoring indicators and clinics assessed were mutually exclusive across the three implementer samples.

To capture quality of care measures in the AHME endline surveys, we used Item Response Theory (IRT) to identify 32 indicators from all 3,178 internal monitoring indicators. IRT has been used extensively to study cognitive and personality traits and to develop computerized adaptive testing (Rasch 1993; Birnbaum 1968). IRT has also become a method for designing, analyzing, scoring, and comparing survey instruments and other tests meant to measure a latent, or unobservable, trait of the respondent (StataCorp 2020). The survey instrument in our study measures the latent trait with a collection of questions of varying difficulty. IRT relates performance on those individual questions and groups of questions to the respondent's level of the latent trait.

To prepare the three samples for IRT logistic models, we transformed ordinal indicators into binary ones and constructed a clinic-level sample for analysis. We adopted a data-driven approach to grouping ordinal levels and selecting one assessment per clinic. Our goal was to maximize data variation in the samples. To this end, MSK indicators were recoded to be at most partially or fully achieved, and SafeCare variables were recoded to be not or at least partially compliant. We used the first assessment from MSK and PSK clinics and the last assessment from SafeCare clinics in our analysis. These choices track well with differences in the structure of monitoring and incentives across organizations. In MSK and PSK clinics, operators faced incentives to improve their indicators over time. As a result, we observed high compliance in later assessments. Although SafeCare clinics faced similar incentives, clinics were eligible to report on more indicators as they progressed through accreditation phases. Thus, we observed very low compliance in the first SafeCare assessment and some improvement over time. The first and last assessment are equivalent for clinics that were assessed only once.

Figure 5.1 presents six criteria that were used successively to reduce the number of indicators. Through this process, the final 32 indicators were identified from the 3,178 provided by implementers. First, we excluded all indicators with data missing from most (50%) clinics. Excluding indicators with high amounts of missing data ensured that the survey instruments would help us learn about quality in most clinics. Implementing partners also did not evaluate clinics on indicators that were unrelated to the services provided by the clinic. For example, all indicators related to inpatient care were excluded in this step. We excluded HIV indicators with data missing from less than 50% of clinics because HIV services are not one of the main AHME program services. In total, this reduced the indicators for consideration from 3,178 to 544 (criterion 1, figure 5.1).

Next, we excluded indicators with no variation (i.e., all clinics in the sample scoring 0 or all scoring 1). This mechanical step is necessary to estimate logistic IRT models that determine further criteria to reduce indicators. In all, 38 indicators lacked variation, thereby bringing the number of candidate indicators to 506 (criterion 2, figure 5.1).

Then, we applied IRT to identify which indicators would best measure quality levels and changes in our samples. Technical and interpersonal quality are examples of latent traits because they are not directly observable but may be captured by other observable items that are monitoring indicators.

IRT models relate such observable items to a latent trait using three main parameters: location, discrimination, and difficulty. *Location* captures the quality level for a given clinic, while the *difficulty* of an item captures the quality level above which 50% of clinics achieve that item. *Discrimination* is the degree to which a survey item can distinguish between relatively low- and high-quality clinics. Importantly, these three parameters allow the quality of a clinic and the difficulty of an item to be measured on the same scale. Formally, we estimate these parameters from binary survey items using logistic IRT models. One-parameter IRT models estimate item-specific difficulty parameters and a common discrimination parameter across all items; two-parameter IRT models estimate item-specific difficulty and item-specific discrimination parameters.

Next, we estimated logistic IRT models and excluded indicators that could not be grouped to (1) estimate a two-parameter logistic model or (2) reject a one-parameter logistic model using a likelihood ratio test at the α=0.10 significance level. First, we considered the official groupings of indicators defined by MSK, PSK, and SafeCare. Then, we relied on face validity (subjective assessment of whether a test captures the content it is supposed to measure) to regroup the indicators for which the two-parameter logistic model did not converge, or we failed to reject the one-parameter logistic model. In total, 174 indicators spanning 12 implementer-specific dimensions of quality (official or customized) met our criteria (criterion 3, figure 5.1). We made one exception for the MSK Clinical Governance grouping, for which we could not reject a one-parameter model (p-value = 0.114). Groupings were mutually exclusive in terms of indicator assignment but not in terms of the latent trait captured, due to similarity in monitoring indicators across implementing partners. Appendix tables 5.A1.1a – 5.A1.1l detail the item-specific difficulty and discrimination estimates from the 12 dimensions of structural quality.

In the next step, we excluded indicators that distinguish poorly between relatively low- and high-quality clinics (that is, indicators with a low discrimination coefficient). We conservatively chose 0.75 as the upper bound of discrimination coefficients for exclusion. We were interested in moderate discrimination coefficients, in addition to high discrimination ones, because those items may detect changes in quality at locations beyond the difficulty level of the item. After excluding indicators with low discrimination coefficients, 125 candidate indicators remained (criterion 4, figure 5.1). Appendix tables 5.A1.2a – 5.A1.2l summarize the results of the two-parameter logistic models re-estimated without low discrimination items for each of the 12 dimensions of structural quality.

Next, we kept four indicators from each dimension of quality that maximized our power to detect a treatment effect. For each combination of four indicators, we estimated the location (that is, quality level) for a representative clinic using the average compliance observed in our samples and the aforementioned IRT models (see appendix tables 5.A1.2a – 5.A1.2l). Then, we increased compliance with each item by 20% and an error term. We repeated this 20 times to create a simulated sample. Finally, we resampled 200 times from the simulated sample to estimate the average treatment effect and bootstrapped standard error. After repeating this simulation for all combinations, we selected the combination of indicators with the largest simulated average treatment effect. This yielded 48 candidate indicators that maximize the power to detect a treatment effect on clinic quality (criterion 5, figure 5.1).

Of the 48 indicators identified by the simulation, 32 were appropriate for the study context and could be feasibly measured in a structured survey (criterion 6, figure 5.1). Specific reasons the other 16 variables were excluded were:

- Not applicable to AHME clinics
- Not applicable to all implementing partners
- Difficult to measure in a structured survey
- An uncommon or universal practice in AHME clinics
- Misaligned with project objectives.

The resulting 32 internal monitoring indicators spanned 13 implementer-specific dimensions of quality (see table 5.2). We incorporated indicators 1–4 into the standardized patient (SP) surveys to capture quality processes (see chapter 6 and table 6.6) and indicators 5–32 into the clinic survey to capture structural quality, in line with the quality of care framework in figure 3.1. Structural quality includes the material resources, human resources, and organizational structure that facilitate the process of seeking and providing care. However, many structural quality indicators could not be incorporated in a single survey question. Rather, a single indicator may be based on an auditor's observation of multiple data points and judgment about them. In these cases, we worked with implementing partners to understand the basis for auditors' judgments and developed survey items that target proxy measure(s) where necessary. In total, we fielded 74 structured questions in the AHME evaluation endline clinic survey to objectively measure the 32 indicators of structural quality. Our analytic sample from the clinic survey included 199 consenting clinics that completed the survey.

5.2 Structural Quality Analysis

We constructed two indices to measure structural quality from the clinic survey. We used clinic responses and enumerator observations from the 74 clinic survey items discussed in section 5.1 to construct 35 binary indicators. We excluded two indicators that were missing data from more than 50% of clinics in our analytic sample. Then, we assigned the remaining 33 indicators to one of two dimensions of structural quality in line with the quality of care framework: (1) structures to facilitate interpersonal care, and (2) structures to facilitate technical care. Interpersonal care captures the exchange of information and preferences between client and provider, and technical care captures the recommendation and implementation of appropriate strategies of care (see figure 3.1).

To evaluate the internal consistency of each set of indicators, we calculated Cronbach's alpha from the analytic clinic sample. Cronbach's alpha relates to variance to inter-item covariance as follows:

$$\alpha = \frac{N\bar{c}}{\bar{v} + (N-1)\bar{c}} \tag{5.1}$$

where N is the number of items, $\bar{c}$ is the average inter-item covariance, and $\bar{v}$ is the average variance.

We calculated compliance scores to summarize each dimension of structural quality at the clinic level. We defined clinic compliance to be the proportion of indicators satisfied at endline. We imputed missing indicator values using clinic compliance with all other observed indicators. We also calculated an overall structural quality compliance measure with the indicators from both dimensions. To visualize the distribution of structural quality compliance, we plotted the kernel density of compliance conditional on treatment status.

To identify the effect of the AHME program on structural quality compliance, we estimated an intent-to-treat (ITT) linear regression:

$$Y_i = \beta_0 + \beta_1 AHME\ Treatment_i + \epsilon_i, \tag{5.2}$$

where Y_i is structural quality compliance, β_0 is the intercept, $AHME\ Treatment_i$ is a binary treatment indicator for whether the clinic was assigned the AHME intervention ($AHME\ Treatment_i = 1$) or was assigned to the control arm ($AHME\ Treatment_i = 0$), and ϵ_i is a normally distributed error term for clinic-level observation i. We assumed the randomization created exchangeable treatment arms to estimate AHME impacts using the ITT parameter. Thus, the coefficient on the AHME Treatment indicator, β_1, is interpreted as the impact of AHME treatment assignment on the structural quality outcome (Y_i). To test the null hypothesis that AHME treatment has no effect or a negative effect (that is, that the coefficient on AHME treatment is less than or equal to zero), we used the Student's t-distribution and report the one-sided p-values. In a sensitivity analysis, we estimated the linear model with each indicator as an outcome (Y_i).

We hypothesized that physical asset capital investments, human capital investments, or provider effort may have driven improvements in structural quality. We tested the first hypothesis using the clinic-reported resale value (in Kenyan shillings, KSH) of investments. Investments were converted to United States dollars (US$, where US$1 = KSH 100), winsorized at the 1st and 99th percentiles, and transformed to hyperbolic arcsine (Bellemare and Wichman 2020). We measured capital investments in the following physical assets:

- Amenities, including air conditioning, room heaters, a water dispenser, and lockers
- Diagnostic equipment, including microscope, glucometer, sphygmomanometer, and ophthalmoscope
- Laboratory equipment, including biochemistry analyzer, ELISA reader, centrifuge, hemogram machine, ultrasound machine, and X-ray machine
- Information technology (IT) equipment, including computer, printer, and projector
- Medical equipment, including beds, exploration tables, negatoscope, oxygen machine, refrigerator, sterilizer, and ventilator machine.

We conceptualized provider effort as performance conditional on ability and estimate effort deficiencies with the "know-do gap"—the difference between provider knowledge and practice-in chapter 6, where we discuss provider competence related to our process quality results. Analyses were conducted in R and STATA MP 14.0 and 15.1 (StataCorp 2020).

5.3 Structural Quality Results

Table 5.3 lists the binary indicators assigned to each dimension of structural quality. Of the 33 indicators, 16 indicators capture structures to facilitate interpersonal care, and 17 indicators capture structures to facilitate technical care. In our sample of 199 clinics, Cronbach's alpha was 0.66 and 0.75 for structures to facilitate interpersonal and technical care, respectively. The social science literature considers an alpha greater than 0.7 to reflect a reliable instrument. Our instrument for structures to facilitate interpersonal care falls short of that rule of thumb. However, Cronbach's alpha for overall structural quality in our sample (0.83) is well above that.

We present the results from our linear models of structural quality compliance in table 5.4. On average, clinics in the control group complied with 52% (SE 0.171) of all structural quality indicators and the AHME program increased compliance by 7.7 percentage points (SE 0.023). The results are remarkably similar for both subdimensions of structural quality. On average, clinics in the control group complied with 51% (SE 0.173) of interpersonal care indicators and 52% (SE 0.204) of technical care indicators; AHME treatment increased compliance by 7.9 (SE 0.024) and 7.5 (SE 0.027) percentage points, respectively. Figures 5.2, panels a–c present the kernel density of structural quality compliance conditional on treatment status. We observe a wide distribution of structural quality compliance among control clinics that narrows slightly among treated clinics. This is most pronounced for overall structural quality compliance.

Figure 5.3 illustrates the AHME treatment coefficients estimated from our models of compliance with 33 individual structural quality indicators. We observe a large range in compliance across indicators: 8.1% of control clinics provided drinking water in the waiting room and recovery room, while 96.7% of control clinics had sufficient natural or electric light. We estimated a positive treatment effect on compliance with most indicators, nine of which are statistically significant at the α=0.10 level with a two-sided p-value. We estimate a large (>0.25) positive effect of the AHME program on the likelihood to comply with both indicators of interpersonal care (informed consent and visible patients' rights chart) and technical care (staff files with credentials and list of referral sites). Among the negative coefficients, all are relatively small in magnitude (>-0.1) and only one coefficient (license to have a private medical practice) is statistically significant at the α=0.10 level using a two-sided p-value.

To explore the mechanisms by which the AHME program improved structural quality, we present the treatment effect on capital investments in table 5.5. The AHME program increased total capital investments by 76% (SE 0.53) relative to the control mean US$2,539. Figure 5.4 presents the kernel density of total capital investments conditional on AHME treatment. We observe a wide, bimodal distribution (i.e., dispersed with two peaks) of structural quality compliance among control and treatment clinics. The increase is due to large increases in investments in IT (coefficient 0.653; SE 0.49) and medical (coefficient 0.866; SE 0.49) equipment. The effects are statistically significant at the α=0.10 and α=0.05 levels, respectively, using a one-sided p-value. The AHME program also increased amenity investments by 17% (SE 0.53), but the effect is not statistically significant. We cannot empirically test for AHME's effects on human capital investments.

5.4 Structural Quality Discussion and Conclusion

We constructed internally consistent measures to capture structural quality and its subcomponents: interpersonal care and technical care. On average, we find low compliance with all dimensions of structural quality in control clinics. The average clinic in the control group complied with 51% percent of interpersonal care indicators and 52% of technical care indicators. However, this masks substantial variation across clinics and individual indicators. Among control clinics, compliance with individual indicators ranged from 8.1% to 96.7%.

We find large, positive effects on compliance with select indicators related to patients' rights and clinical record keeping. The AHME program increased compliance with interpersonal care indicators by 7.9 percentage points (SE 0.024) and technical care indicators by 7.5 percentage points (SE 0.027). The AHME program significantly increased the likelihood that a clinic would offer informed consent, display a patients' rights chart, keep staff files with credentials, and keep a list of referral sites by more than 25 percentage points. Improvements in structures to facilitate technical care and interpersonal care may have been driven by capital investments in amenities, IT, or medical equipment. However, measurement of the value of investments is noisy, and we lack the power to detect a statistically significant effect of the AHME program after correcting for multiple hypothesis testing. Moreover, we cannot rule out the possibility of AHME investment effects in human capital. In chapter 6, we discuss the know-do gap as a proxy for provider effort, but next we turn to process quality.

Chapter 6

Process Quality

This chapter focuses on the second type of quality of care from our framework: process quality. Specifically, we seek to answer the following three research questions related to process quality:

1. To what extent did the AHME program have an impact on correct actions related to process quality?
2. What was the effect of the AHME program on nonessential or unnecessary actions related to process quality?
3. What is the effect of the AHME program for different services and clients with different characteristics?

After discussing our process quality methodology, we report the effects of AHME on technical process quality components. Overall, we find that the AHME program significantly reduced the likelihood that outpatient clients received correct actions defined as minimal and essential. Specifically, the AHME program resulted in a significant 12% reduction of correct case management compared to clinics that did not receive the program. Unnecessary diagnostic and treatment actions were also lower among AHME treatment clinics versus control clinics, suggesting reduced waste related to unnecessary care. Despite how these results were not significant, their reductions were consistent in magnitude with the drop in correct care.

Regarding whether providers perform integrated management of childhood illnesses (IMCI) actions identified in chapter 5, our unannounced SP visits revealed that such actions were almost never carried out neither inside nor outside of the program. Yet we observe high rates for these indicators in the program data. Therefore, we caution against the use of direct observation as a method for implementers to collect quality measures, because they can be subject to biases from the Hawthorne effect (when individuals change their behaviors because they are being watched) and social desirability (the tendency of survey respondents to answer questions in a manner that will be viewed favorably by others).

In both AHME treatment and control clinics, we found alarming deficits in laboratory quality (outside the scope of the AHME program), with follow-on treatment that can result in individual and public health consequences. Also, in both treatment and control clinics, we found poor clients received significantly lower rates of both correct and unnecessary actions compared to the non-poor.

In this study context, SPs excelled at revealing what occurs when the same patient visits a random subset of the countrywide clinic sample, presenting the same condition as a new walk-in patient. By experimentally varying SP case presentation, we were able to identify the effects of certain client characteristics (age, gender, demanding medicines, marital status, and poverty) on process quality measures. We find that with each additional year of age, individuals (between the ages of 22 and 35 - this is the actual age range of the SPs at the time of fieldwork) were more likely to receive unnecessary lab tests, but not medicines. Females in general were more likely to receive non-efficacious medicines, but were as likely to receive lab tests (any unnecessary lab tests, total lab tests, and total unnecessary lab tests) than males. Those who demanded a non-efficacious drug got more correct care; the poor got fewer instances of correct and nonessential care; and there were no observed differences between unmarried and married clients presenting for family planning services. However, this part of the study cannot show or reflect the patient mix and patient sorting that occurs in reality. The remainder of the chapter details process quality measurement and analysis, results, and discussion.

6.1 Process Quality Measurement and Analysis

To assess process quality of care in low- and middle-income country settings, the literature outlines several methodologies, including direct observation, administrative or medical record abstraction, client exit interviews, provider vignettes, and standardized patients (SPs). Each method has its own interpretation and set of advantages and disadvantages (Peabody et al. 2000; Das, Hammer, and Leonard 2008; Kwan et al. 2018).

To answer the AHME impact evaluation research questions related to process quality, we implemented two main survey methodologies: SPs and provider vignettes, which minimize bias in assessing provider practice and provider knowledge, respectively. Other methods to assess quality of care have certain limitations that do not make them ideal for answering our research questions. Health and medical record data often do not exist in low- and middle-income country settings and when they do, they suffer from bad data quality. Direct observation is biased by the Hawthorne Effect. Client exit interviews represent different patient sorting and patient mixes across clinics, and not only do clients not always understand medical jargon, but it is difficult to know precisely what medical condition the client has. Vignettes excel at assessing provider knowledge, but as for practice, vignettes are subject to social desirability bias and differ largely from practice measures (Kwan et al. 2019). For example, the “know-do gap” is a well-documented phenomenon in the literature referring to the difference between provider knowledge and provider actions in practice (Das et al. 2015; Mohanan et al. 2015).

Because of the ability for unannounced SPs to mimic an actual client-provider interaction, the SP method, which is often used for medical training in high-income countries, is increasingly considered the gold standard for measuring provider practice in low- and middle-income settings. The SP method is not without limitations as it requires a one-time visit for services that (1) do not subject the client to invasive procedures; (2) can only assess tracer health conditions that have been validated for ethical research; and (3) do not require established client services or follow-up visits, such as those related to chronic conditions or other ailments. However, we identify the method's limitations as favorable conditions to assess the quality of walk-in outpatient services of interest in the AHME program. In this study, we further combine the use of SPs and vignettes to understand provider practice relative to provider knowledge.

Data Sources

Standardized Patients

The SP data reflects process quality measures for clients seeking walk-in, outpatient services. The SP data come from 1195 successful SP visits to clinics in the AHME evaluation clinic sample. All SP visits were conducted by trained actors between February and May 2019 at 211 clinics. To ensure accurate and comprehensive recall, within 1 to 3 hours after each SP visit, SPs completed an exit questionnaire administered by a fieldwork supervisor.[1] The exit questionnaire collected information regarding the SP's visit, including time of arrival, time of departure, history questions asked, diagnosis, lab tests ordered, medicines dispensed and prescribed, counseling given, and a subjective assessment of the visit. Further, for each visit, SPs and their field supervisors attempted to identify all providers seen by the SPs. The list of providers formed the provider survey sampling frame.

In January 2019, 40 individuals were locally recruited, trained, and hired as SPs portraying one of four case scenarios: childhood diarrhea with a child in absentia, family planning, adult mild asthma, or adult malaria.[2] None of the trained actors possessed the conditions described in the scripted case scenarios, and all were seemingly healthy so providers would not detect and treat other health ailments that were unrelated to the assigned case scenario. In that manner, all individuals were standardized in presentation to portray the SP case scenarios for this study.[3]

[1]These debriefs initially occurred in person; however, as the field team moved to areas that were harder to access and more remote, some debriefs occurred over phone within 1 to 3 hours.

[2]We implemented our SP study based on protocol from previously implemented SP studies in China, India, and Nairobi, the capital of Kenya (Sylvia et al. 2014; Das et al. 2012; Daniels et al. 2017; Kwan et al. 2019), and the four cases were adapted from other SP studies in Senegal, Kenya, and Tanzania (Daniels et al. 2017; Goodman et al. 2017).

[3]See appendix 6.A1.1 for additional details for SP case development, recruitment, training, and piloting. See appendix 6.A1.2 for ethical considerations of the SP method in the AHME evaluation. See appendix 6.A1.3 for the granted request for a waiver of informed provider consent, conditional on disclosure of SP study participation to any clinics receiving SPs. See appendix 6.A1.4 for the SP script and exit questionnaire

To identify the effect of specific client characteristics on quality of care, we developed experimental variants for each of the SP cases and randomly assigned them to the clinics in the sample. An SP audit study that experimentally varies one characteristic is the standard method to identify this effect. In our study, we implemented the following experimental case variants to assess effects of client characteristics on care: demanding vs. not demanding a non-efficacious and harmless drug (albendazole) for childhood diarrhea; demanding vs. not demanding a non-efficacious and harmful drug (amoxicillin) for the childhood diarrhea SP case; unmarried vs. married for the family planning SP case; not being able to afford more than 300 Kenyan Shillings (KSH, where US$, where KSH 100 = 1 United States Dollar, US$) vs. affording clinic fees for asthma and malaria SP cases; and female vs. male for asthma and malaria SP cases.

Each of these experimental case variants was developed for different cases based on feasibility and was randomly assigned to the clinics during sampling with statistics software.[4] For asthma and malaria, we refer to the experimental case variant in which the SP states he or she cannot afford more than 300 KSH as a proxy for a "poor" client coming for services, using US$3.00 as an approximate benchmark for the World Bank lower-middle income class poverty line value of US$3.20, adjusted for purchasing power parity (World Bank Group, 2020 (accessed April 21, 2020)). Table 6.1 includes a description of all four case scenarios, opening statements, and experiments conducted by case.

The SP data consist of 1195 SP-provider interactions, where each interaction is a successful SP visit. We define successful visits as visits where the SP visited the clinic during operating hours and interacted with clinic staff, similar to an actual client presenting with similar conditions. Thus, the 1195 SP interactions include visits where the SP was referred or turned away after presenting an opening statement to staff at the clinic: for example, because of presenting with the child in absentia or not having enough money. The data do not include visits where the clinic was closed permanently.[5] See figure 6.1 for the SP survey

for childhood diarrhea case scenario. Additional scripts, exit questionnaires, and supervisor questionnaire for the SP cases and experiments are available upon request.

[4] For the childhood diarrhea case, half of the clinics were randomly assigned to receive an SP demanding a non-efficacious, unharmful medicine for the presenting condition (albendazole, which is used for deworming) with the remaining half to receive an SP demanding a non-efficacious, harmful medicine for the presenting condition (amoxicillin, an antibiotic that treats bacterial infections). Demanding SPs were trained to demand the assigned medicine at the very end of the visit, and we captured actions before the SP demands (pre-demanding) and after the SP demands (demanding) to assess the effects of demanding. For the family planning case, half of the clinics were randomly assigned to receive the unmarried family planning SP, with the remaining clinics to receive the married family planning SP. For the asthma and malaria cases, half of the SPs were female and the other half were male. Supervisors and SPs were not informed of our research questions related to SP gender, and we assume that male and female SPs were assigned "as good as random" to the clinics, following Daniels et al. (2019). For the asthma case, one-third of the clinics were assigned to receive the poor SP case, and the data for the remaining clinics reflected a not-poor SP case. For the malaria case, all clinics received both the poor SP case and the not-poor SP case.

[5] If clinics were temporarily closed, the SP field team determined when the clinic would be open and would plan another visit accordingly. If a clinic remained closed after three attempts, we captured the reason and considered the visit unsuccessful. Unsuccessful visits are not included in the 1195 interactions.

sampling flowchart (a graphical depiction of the evaluation clinics, assignment of cases and case experiments for the SP surveys).

A technical advisory group[6] 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.

Provider Survey and Vignettes

The provider survey captures data on provider characteristics, knowledge measures, and provider preferences. During November and December 2019, we implemented a provider survey at 199 of the 232 clinics in the evaluation sample. Respondents were providers seen by SPs. If providers seen by SPs were unavailable, no longer worked at the clinic, or were not identified in SP fieldwork, other respondents who provided outpatient services at the time of interview for family planning, adult curative, and child curative services were interviewed and served as replacements in the provider analytic sample. Overall, 322 of 344 providers identified during SP fieldwork were successfully interviewed at the 199 clinics in the AHME evaluation clinic survey sample.

The provider survey measured respondents' age, qualification, and other characteristics. The provider survey also contained two modules to assess provider knowledge and provider preferences (discussed in chapter 8). To assess provider knowledge, respondents were asked to complete three provider vignette cases, where provider knowledge was assessed for childhood diarrhea, family planning, and malaria cases with similar presentation as the SP cases. At 178 of the 199 clinics in the clinic evaluation sample, 288 providers completed the childhood diarrhea vignette, 285 providers completed the family planning vignette, and 287 providers completed the malaria vignette. Figure 6.1 depicts the provider survey and vignettes samples and how they link to the SP surveys. Appendix 6.A1.5 includes more details on the provider vignettes.

Process Quality Measures

Main Outcomes: Correct and Unnecessary Care

With the 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 6.2. Correct case management definitions refer to minimal and essential actions as benchmarked against national guidelines for case management (Kenya Ministry of Health 2010b; Kenya Ministry of Public Health and Sanitation 2010; Kenya Ministry of Health 2010a, 2014; Kenya President's Malaria

[6] Additional information on the role of the technical advisory group is included in appendix 6.A1.1.

Initiative 2018). We confirmed these definitions for minimal, essential actions, as well as our classifications for valid or unnecessary lab tests and medicines, with guidance from our technical advisory group.

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

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

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

- ***Management of Malaria:*** For malaria, providers were coded as correctly managing the case if they ordered a malaria rapid diagnostic test (RDT or mRDT) or malaria microscopy test, and not if otherwise.[7] 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.

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

Other Case-specific Outcomes

We also examined interpersonal and other technical process quality outcomes for each SP case. These are clinically relevant actions for each scenario, such as history questions asked, diagnostic and treatment actions, counseling given, and whether the provider asked the SP to return or referred the SP elsewhere for further management.

Family planning has a different visit structure compared to the other SP cases, which are designed to seek child and adult curative services. Instead of an ailment with a diagnosis and specific testing and treatment recommendations, the family planning SP scenario involves interpersonal process measures, such as counseling a client on family planning procedures until she has reached a preference. We examine whether the provider mentioned side effects for modern family planning methods, provided an explanation of the effectiveness of traditional methods, and recommended modern and traditional family planning methods to the client.

In this study, we assess quality of child and adult curative services with more technical process measures, along with some interpersonal process measures. For malaria, national malaria recommendations in Kenya advise providers to prescribe AL as a first-line medication and paracetamol as a cost-effective option (preferable to ibuprofen, for instance) to manage fever symptoms. In line with these recommendations, we examined diagnostic actions (that is, taking the client's temperature, ordering a malaria test); whether diagnostic test results were positive; and whether the provider dispensed or prescribed AL and paracetamol. Then, we examined the extent the SP's demand for a harmless drug (albendazole) or a harmful drug (amoxicillin) had an effect on providers' dispensing/prescribing rates of non-efficacious medicines for the childhood diarrhea case. We are able to identify this effect by randomly assigning SPs to demand the harmless or harmful medicine at clinics.

Estimating Equations

To evaluate the impact of the AHME intervention on process quality, we estimated intention to treat (ITT) effects using linear regression models, with standard errors clustered at the clinic level. These analyses rely on randomization for an unbiased estimate of the impact of the AHME program. We note deviations from the format of this estimating equation in the sections that follow. To understand the impact AHME had on process quality, analyses using SP data were conducted at the SP-provider visit level; analyses using client exit interview data were conducted at the client level; analyses using the household survey were conducted at the individual level, where multiple members of a single household were included; and analyses using provider survey data only were conducted at the provider level. When SP data were linked to provider survey data, one observation was considered a successful SP-provider visit with provider survey responses from the provider seen during the SP visit. For analyses at the level of the SP-provider visit, we also included SP case scenario and SP actor fixed effects.

For SP analyses, we made two assumptions using ITT parameters: (1) the randomization of AHME clinic assignment created exchangeable AHME treatment arms to estimate AHME impacts; and (2) the randomization of SP experiment assignments created exchangeable SP treatment arms to estimate the impacts of SP experiments. We used the following regression (equation 6.1) to analyze the SP data:

$$Y_i = \alpha + \beta AHME\ Treatment_i + \sum_{(v=1)}^{V} \gamma_v CaseVariant_i + CaseFE_i + SPIndividualFE_i + \epsilon_i, \tag{6.1}$$

where the α is the intercept, and $AHME\ Treatment_i$ is a binary treatment indicator for whether the clinic associated with the SP-provider visit received the AHME intervention ($AHME\ Treatment_i = 1$) or was assigned to the control arm ($AHME\ Treatment_i = 0$). $CaseVariant_i$ is a set of V indicators for each of the SP experimental case variants, where each variant is identified at the SP case level. $CaseFE_i$ are SP case fixed effects, and $SPIndividualFE_i$ are SP actor fixed effects. The coefficient on the $AHME\ Treatment_i$ indicator variable, β, is interpreted as the impact of the AHME intervention on the outcome of interest (Y_i). The coefficients on each of the SP experiment indicator variables, γ_v , where v takes on integer values 1 to V corresponding to each of the SP experiments, are interpreted as the effect of the client characteristic corresponding to the SP case variant on the outcome of interest (Y_i); and ϵ_i is a normally distributed error term for observation i, clustered at the clinic level.

For our main outcomes of interest, SP analyses were conducted on data containing all SP-provider visits for each case separately ("case-wise" separately for diarrhea, family planning, asthma, and malaria) and for all cases pooled ("pooled", n = 1195 interactions) with SP case fixed effects. To test whether the effect of AHME on any of our outcomes was different for different client characteristics (that is, the values representing our SP experiments), for each SP model we presented, we conducted additional specifications to account for possible mediating effects between AHME treatment and each of the SP experiments. For these analyses we used conditional linear regression that included additional interaction terms that could impact the outcome of interest (for details on methods, see appendix 6.A1.1). The results are noted in the next section and found in the appendices to chapter 6.

For analyses at the provider level, we used the following conditional linear regression (equation 6.2), including additional covariates that could impact the outcome of interest:

$$Y_i = \beta_0 + \beta_1 AHME\ Treatment_i + \beta_2 X_i + \epsilon_i, \tag{6.2}$$

where β_0 is the intercept and $AHME\ Treatment_i$ is a binary treatment indicator for whether the clinic associated with the individual received AHME ($AHME\ Treatment_i = 1$) or was assigned to the control arm ($AHME\ Treatment_i = 0$). The coefficient on the AHME treatment indicator variable, β_1, is interpreted as the impact of the AHME intervention on the outcome of interest (Y_i), and ϵ_i is a normally distributed error term for observation i, clustered at the clinic level. Covariate X_i is added to control for potential baseline imbalances, and β_1 is interpreted as the impact of $AHME\ Treatment_i$, conditional on X_i.

All p-values reported in this chapter are two-sided, unless stated otherwise. All analyses were conducted in R and STATA MP 14.0 and 15.1 (StataCorp 2020).

6.2 Process Quality Results

This results section begins with summary statistics of the SP data, followed by results of the AHME program's effect on technical process quality measures in terms of correct actions (correct case management and AHME internal monitoring indicators) and unnecessary actions (any unnecessary lab tests and any unnecessary medicines). We conclude this results section with case-specific findings across each of the different case scenarios and then describe how specific client characteristics play a role in the quality of care received.

Summary Statistics

First, we highlight summary statistics from the SP data. Table 6.4 reports summary statistics for the SP data, including summary statistics on lab tests and medicines. First, based on SP reports, 18% of providers for the 1195 SP-provider visits appeared to be less than 30 years old; 58% were aged 30–50 years; and 25% were older than 50. In all, 39% of providers were female (see "Pooled" column in table 6.4). The average number of clients in the waiting room was 1.5 for the 1195 SP interactions. SPs spent an average of 6.5 minutes with the provider and were asked an average of 5.9 history questions. SPs were asked to return in 40% of visits, and fewer than 3% of the interactions resulted in a referral.

In terms of provider practice and knowledge, we find that on average, 60% of the pooled SP data resulted in correct case management. Although providers of 90% of the diarrhea SP interactions knew how to manage a diarrhea case, they managed only 70% of the diarrhea SP interactions correctly. We find knowledge to be particularly low for correctly managing the family planning case: 11% of the providers knew how to manage the case, although 25% did so in practice. This disparity 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 performs at least one of the four actions, we find that knowledge and practice are similar and high. Providers who saw the family planning SPs knew to do at least one of the four actions for 94% of the interactions; and in practice, the same providers performed at least one of the four actions for 94% of the interactions (not shown). For malaria, 79% of malaria SP interactions resulted in correct case management, despite the providers of 98% of malaria SP interactions knowing how to correctly manage the case. Because we did not administer a clinical vignette for asthma, we do not know how knowledge differs from practice specifically for this case.

Multivariate Regression Results: Correct and Unnecessary Actions

Next, we turn to ITT results from multivariate regression analyses on the SP data. Overall, we find that the AHME program significantly reduced the likelihood that outpatient clients received correct actions defined as minimal and essential, and we observe a reduction in unnecessary actions, though this is not significant.

Correct case management

Beginning with correct actions, table 6.5 reports the effect of the AHME program on correct case management from SP data for diarrhea (column 1), family planning (column 2), asthma (column 3), malaria (column 4), and all cases pooled (column 5).[8] We see that AHME treatment assignment decreased the likelihood of correct case management by 7.7 percentage points (p-value = 0.021; model 5) on average, compared to the control group (mean = 63.6%, a relative reduction of 12%). When examining the different cases separately (models 1–4), we see that the coefficient on the AHME treatment indicator is significant only for the malaria scenario (coefficient = -0.092, p-value = 0.058). However, the coefficient on the AHME treatment is large, negative, and similar for each case, and pooling increases statistical power to detect the significant reduction in correct case management due to the AHME program.

Internal monitoring indicators: Childhood diarrhea case

Regarding the AHME program's internal monitoring indicators identified in chapter 5, we examine the four process quality indicators related to the Integrated Management of Childhood Illnesses (IMCI) approach (table 5.2, indicators 1-4). We use the childhood diarrhea SP data (n=189 pre-demanding and n=189 post-demanding observations for 189 successful and completed visits at 189 clinics), where unannounced SPs visited clinics reportedly with a child at home sick with diarrhea. We find no evidence that the AHME program had an effect on any of these four indicators at the α=0.10 significance level (table 6.6).[9] In fact, we see that the unannounced SP visits revealed all four indicators in both AHME treatment and control clinics were almost never performed: whether the provider asked if the child has cough or difficulty/rapid breathing was asked in 6.3% of all clinics; 0.5% of the clinics resulted in the provider asking the SP about difficulty/rapid breathing; no clinics had a

[8] Appendix Table 6.A2.1 reports the model shown in table 6.5 with interactions between AHME treatment and the SP experiments.

[9] See appendix table 6.A2.2b for regression models including the interaction between AHME treatment and the SP experiment for childhood diarrhea case.

provider ask if anyone in the family was coughing; and 0.5% of the clinics had the provider ask about the child's HIV status (appendix table 6.A2.2a).[10]

Next, we turn to whether the AHME program assignment had effects on providers ordering any unnecessary (not valid) lab tests and any unnecessary (non-efficacious) medicines. Not only do we find a significant reduction for correct case management, we also find that AHME treatment clinics are less likely to order any unnecessary lab tests (coefficient = -0.019 or 12.3% reduction, p-value = 0.423; table 6.7, model 5) and less likely to order any non-efficacious medicines (coefficient = -0.050 or 8.5% reduction, p-value = 0.139; table 6.8, model 5).[11] However, the reductions in unnecessary actions in the pooled SP data are not statistically significant at α=0.10.

Unnecessary tests

Examining any unnecessary tests by case, we find heterogeneous results. First, we find a significant 51% reduction (coefficient = -0.070, p-value = 0.015) for the diarrhea SP case in AHME treatment clinics from the control group mean 0.138 (table 6.7, model 1). The control group ordered low rates of any unnecessary tests for family planning (mean = 0.022, table 6.7, model 2) and asthma (mean = 0.065; table 6.7, model 3), with negative effects from the AHME program assignment that were not significant at the α=0.10 level. The control group mean for any unnecessary lab tests for the malaria scenario was 0.289, with a 15% nonsignificant increase (p-value = 0.415) among the AHME treatment group (table 6.7, model 4).

Non-efficacious medicines

As for non-efficacious medicines, we find AHME treatment significantly reduced whether any non-efficacious medicines were ordered for the diarrhea case (coefficient = -0.106, p-value = 0.027) compared to AHME control group mean = 0.723 (table 6.8, model 1), but the AHME program assignment did not affect the other cases at the α=0.10 significance level (table 6.8,

[10] We also examined the PSK administrative data for each of the measures. We find much higher rates in the data compared to our findings: (1) Provider asked if child was coughing: 83.8% of 37 clinics at first observation and 82.6% of 23 clinics at second observation; (2) Provider asks if there is difficulty in breathing, fast breathing, wheezing: 67.6% of 37 PSK clinics at first observation and 82.6% of 23 PSK clinics at second observation; (3) Provider asks whether the mother or other family members are coughing: 21.6% of 37 PSK clinics at first observation and 47.8% of 23 PSK clinics at second observation; (4) Provider ask for HIV positivity in mother or child: 27.0% of 37 PSK clinics at first observation and 39.1% of 23 PSK clinics at second observation. The PSK administrative data were collected with direct observation, where an auditor directly observed actual client-provider interactions. This differs from our data, which are measured by unannounced visits conducted by an actor trained in a standardized patient scenario. We believe the administrative data observations conducted by auditors associated with the PSK program may be subject to both the Hawthorne Effect and social desirability bias.

[11] See appendix tables 6.A2.3 and 6.A2.4 for regression models on any unnecessary lab tests and any unnecessary medicines outcomes, respectively, including interactions between AHME treatment and each SP experiment.

models 2–4). Among all clinics, we find low rates of any unnecessary medicines for family planning (table 6.8, model 2).

Multivariate Regression Results: Process Quality by Service

Next, we assess the impact of AHME on process quality by service type. First, we start with the health care service of family planning, which relies heavily on counseling the client so that she can choose a family planning option appropriate for her. Then, we present results for the asthma and malaria cases. We examine appropriate and inappropriate diagnostic actions, rates of laboratory testing and results, and treatment for these two adult curative services. Lastly, we report our findings for the childhood diarrhea case. The services we examine with SP cases comprehensively highlight different interpersonal and technical components of health care processes that may occur in many patient-provider interactions.

Preventive Health Care Services: Family Planning

For family planning, we do not see any significant program effects on any of the following individual components of correct case management (table 6.9, panel a): asked family planning history, asked obstetric history, or asked which family planning method was preferred (models 1, 2, and 4). We do find a 32.3% reduction (coefficient = -0.168, p-value = 0.024; model 3) of ruling out pregnancy among AHME treatment clinics from the AHME control group mean 0.505.[12] This is driven by asking whether the female is pregnant, rather than through a pregnancy test (results not shown).

We do not find any statistically significant program effects related to modern family planning methods at the α=0.10 level (table 6.9, panel b, models 1 and 2), but do find program effects related to traditional family planning methods.[13] AHME treatment clinics were 36% more likely to mention the explanation of effectiveness for traditional methods (coefficient = 0.075, p-value = 0.060) compared to the AHME control group mean 0.055 (table 6.9, panel b, model 3). AHME treatment clinics were 59.4% less likely to suggest a traditional method (coefficient = -0.098, p-value = 0.027) compared to the AHME control group (table 6.9, panel b, model 4).

Adult Curative Services: Asthma and Malaria

For adult curative services, we do not find significant program assignment effects on correct case management, whether any lab tests were ordered, or whether non-efficacious medicines were given for asthma (see model 3 in tables 6.5, 6.7, 6.8), but we find program effects related to diagnostic and treatment actions for malaria.

[12] See appendix table 6.A2.5a for the model for family planning correct management components including an interaction between the unmarried SP experiment and AHME treatment.

[13] See appendix table 6.A2.5b for the model for family planning counseling on modern and traditional family planning methods including an interaction between the unmarried SP experiment and AHME treatment.

If prompted by the provider, the malaria SP is designed to mention a fever "that comes and goes." We first assess whether providers assess body temperature. We find higher rates of taking the temperature with a thermometer (AHME control group mean = 0.494) than by touch (AHME control group mean = 0.060) among all malaria SP data (table 6.10, panel a, models 1 and 2). Interestingly, we find a nonsignificant 8.7% reduction of taking the temperature with a thermometer at the α=0.10 level and an 83% increase (p-value = 0.091) for taking temperature by touch due to AHME (see table 6.10, panel a, models 1 and 2), suggesting a substitution effect among AHME treatment clinics.

In this study, providers were categorized as correctly managing the malaria SP if they ordered any malaria diagnostic test—either mRDT or microscopy. Among the malaria SPs who had a complete visit (n=372 out of 392), we find the AHME program reduced whether the provider offered at least one malaria diagnostic test by 9.6% (coefficient = -0.084, p-value = 0.085) relative to the AHME control group mean 0.875 (table 6.10, panel a, model 3).[14,15] Further, the policy in Kenya for malaria management warns of "high false positive rates for mRDT" so providers should conduct "microscopy with high quality, functioning labs" (Ministry of Public Health and Sanitation 2010b; President's Malaria Initiative 2018). In a multivariate regression, we do not observe differences between AHME treatment vs. control clinics on either the use of RDT or the use of microscopy at the α=0.10 significance level (table 6.10, panel a, models 4 and 5).

Despite high rates of correctly ordering any malaria test, we find suboptimal laboratory quality in both treatment and control groups. Alarmingly, 18.0% of 122 malaria RDT tests and 33.3% of 192 malaria microscopy tests given to SPs were positive (results not shown). Since all SPs tested negative with malaria microscopy tests at a known high-quality laboratory at the beginning and end of 22 days of malaria SP fieldwork. This was (1) part of research protocol and (2) for eligibility to present as an actor for this case, we consider 18.0% and 33.3% to be true false positive rates of mRDT and malaria microscopy in the study sample. We find no difference between AHME treatment and control clinics for whether the SP is malaria positive unconditional on receiving a test (table 6.10, panel b, model 1). However, among AHME treatment clinics, we find a 50.5% increase (coefficient = 0.110, p-value = 0.042) compared to the AHME control group mean 0.218 (table 6.10, panel b, model 2) in whether the SP tests malaria positive conditional on receiving a test.

The AHME program had no effect (coefficient = 0.018, p-value = 0.839) on mRDT positive test results when compared to the control group (mean = 0.177; table 6.10, panel b, model 3). However, malaria microscopy positive test results were 51.3% higher (coefficient = 0.148, p-value = 0.088) in the AHME treatment group relative to the control group mean 0.288 (table 6.10, panel b, model 4), despite no statistical difference in proportions of ordering a microscopy test between the AHME treatment and control clinics (0.512 vs.

[14]By examining all 392 successful malaria SP visits, we find 85% were offered at least one malaria diagnostic test in control group clinics, with a 10.8% lower rate in AHME treatment groups (coefficient = -0.092, p-value = 0.058; table 6.5, model 4)

[15]Only 3 malaria SP visits resulted in the provider ordering both malaria diagnostic tests.

0.500, t-statistic = 0.23, two-sided p-value = 0.819).[16]

For medicines, we find that AHME treatment clinics had slightly higher rates of dispensing or prescribing artemether lumefantrine (AL) than AHME control clinics (AHME control group mean = 0.269; table 6.10, panel b, model 5), but this was not significant (coefficient = 0.062, p-value = 0.244). AHME treatment clinics dispensed and prescribed paracetamol at lower rates than the control group (35.8% reduction, p-value = 0.017 from AHME control group mean = 0.237, table 6.10, panel b, model 6).

We find significantly different proportions of dispensing/prescribing AL for SPs who were malaria positive vs. negative (0.655 vs. 0.202, t-statistic = 8.68, two-sided p-value < 0.001). Among 84 visits where the SP was malaria positive unconditional on receiving a test, 65.5% (n=55) of SPs were prescribed/received AL; whereas, among 287 visits where the SP was malaria negative unconditional on receiving a test, 20.2% (n=58) of SPs were prescribed/received AL. We see that having a (false) positive test result increases the rate of receiving first-line malaria treatment, both conditional on receiving a test and unconditional on receiving a test. When clients test positive for mRDT or malaria microscopy, national guidelines recommend AL as the first-line malaria treatment; however, given the high false positive rates, onward treatment for true negative individuals who test positive is worrisome and can have individual and public health consequences. It is additionally worrisome that test negative individuals are dispensed or receive AL.

Child Curative Services: Childhood Diarrhea

Table 6.11 reports results for dispensing/prescribing medicine for the childhood diarrhea SP case.[17] We find no significant program effects on whether providers dispensed or prescribed the correct treatment (ORS or zinc) at the α=0.10 significance level (table 6.11, models 1 and 2). However, AHME treatment clinics were 6.4 percentage points less likely (14.4% reduction, p-value = 0.031) to prescribe or dispense any non-efficacious medicines when compared to the control group mean 0.723 (table 6.11, model 3). However, we do not find a difference between AHME treatment and control groups on dispensing/prescribing the two non-efficacious medicines targeted for SP experiments at the α=0.10 significance level (table 6.11, models 4 and 5).

Multivariate Regression Results: Process Quality by Patient Characteristics

Next, we turn to our results on the effects of quality of care on different client characteristics: gender, age, demanding a medicine, marital status, and poverty. We assess whether AHME

[16] See also the AHME coefficient (0.017, p-value = 0.790) on multivariate regression model 5 in table 6.10, panel a.

[17] See appendix table 6.A2.7 for childhood diarrhea treatment regressions with an interaction between the demanding SP experiment and AHME treatment.

had different effects for different characteristics and whether different client characteristics had an effect on process quality at the α=0.05 significance level.

For age and gender, we assessed SP actor characteristics (see appendix table 6.A2.8) as covariates in multivariate regression models and removed SP fixed effects from our standard model (equation 6.1). For marital status, wealth, and demanding a medicine, we examined the randomly assigned SP experiments as covariates in the model for equation 6.1. For each client characteristic, we visually summarized the effects by plotting the coefficient and standard errors and adjusted for multiple hypotheses with Benjamini and Hochberg (1995) procedures. We pooled the asthma and malaria SP data to examine the effects of SP gender and SP age, shown in figures 6.2 and 6.3, respectively. The effects of demanding medicine, being unmarried, and being poor are shown in figure 6.4 for the diarrhea SP case, figure 6.5 for the family planning case, figure 6.6 for the asthma SP case, and figure 6.7 for the malaria case, respectively.

Gender

First, among the asthma and malaria SP data, we find no difference between female and male SPs in receiving correct case management (coefficient = 0.010, p-value = 0.800; appendix table 6.A2.9a, model 1).[18] We also do not see any gender effects for laboratory tests; however, we observe gender effects on unnecessary medicines (figure 6.2). Compared to males, female SPs are more likely to receive any non-efficacious medicines (coefficient = 0.090, p-value = 0.036) and a higher number of medicines (coefficient = 0.314, p-value = 0.013). We find evidence that females receive a higher number of non-efficacious medicines (coefficient = 0.394, p-value < 0.001), but there is no difference in the number of efficacious medicines received by males versus females (appendix table 6.A2.9a, models 3, 6, 7, 8).

Age

We find no difference in correct case management across SP age (appendix table 6.A2.9a, model 1).[19] However, in contrast to SP gender effects, we find age effects on unnecessary laboratory testing, but not medicines (figure 6.3). SP actors ranged in age from 22 to 35 years. They portrayed the asthma SP case at 24 (females) or 25 (males) years of age, and the malaria SP case at 28 years of age. With each additional year of actual age, despite portraying a fixed age by case (24/25 for asthma; 28 for malaria), SPs were more likely to receive any unnecessary lab tests, more lab tests, and more unnecessary lab tests (see appendix table 6.A2.9a, models 2–8). However, we find that AHME treatment clinics gave significantly fewer medicines and fewer non-efficacious medicines to SPs with each additional year of actual age in the 22–35 age range (appendix table 6.A2.9b, models 6 and 8).

[18] This is robust to different specifications, including interactions between the AHME treatment indicator and SP age (appendix 6.A2.9b, model 1).

[19] See appendix tables 6.A2.14b, 6.A2.15a, and 6.A2.15b for different specifications, including SP characteristics and AHME treatment interactions without and including provider characteristics.

Effects of demanding a medicine

We find those who demand a non-efficacious drug are significantly more likely to receive correct care and both components of correct care (gave or advised on ORS and asked SP to return or referred SP) at the α=0.05 significance level. We find no differences in the likelihood of receiving any lab tests or any unnecessary lab tests between demanding and pre-demanding at the α=0.05 significance level (see figure 6.4). Interestingly, we find that demanding an unnecessary, harmless drug for a childhood diarrhea case significantly increases its rate of dispensing by 165.1% (coefficient = 0.137, p-value = 0.004; table 6.11, model 4); however, demanding an unnecessary, harmful drug for this case did not significantly increase its rate of dispensing (coefficient = 0.025, p-value = 0.527; table 6.11, model 5). We did not find evidence for AHME treatment and comparison clinics treating demanding and pre-demanding SP differently (see appendix table 6.A2.7). Based on further inspection, we find providers who are assigned the harmful amoxicillin experiment are significantly more likely to advise or dispense/prescribe ORS, and providers who are assigned the harmless demanding experiment are not more likely to advise or dispense/prescribe ORS.[20]

Marital status

We find that providers were less likely to rule out pregnancy at the α=0.05 significance level (coefficient = -0.176, p-value = 0.041; table 6.9, panel a, model 3) in SP interactions in which SPs portrayed being unmarried versus being married. Otherwise, we find no differences observed by marital status at the α=0.05 significance level when adjusting for multiple hypotheses using Benjamini and Hochberg (1995) procedures (see figure 6.5). However, imprecise estimates may be due to a lack of statistical power. There were no differences between AHME treatment and comparison clinics on the effects of being unmarried.[21]

[20] We have scrutinized a few possible channels in our data that are related to either provider behavior or limitations of the method we implemented. Based on the survey design, the increase in correct care for diarrhea can only be related to variables where we have captured the outcome at two time points: pre- and post-demanding. Thus, effects of demanding on correct care is related to advising on ORS or referral/return. It cannot be from dispensing/prescribing ORS, which is captured once at the end of the visit. One can imagine that having two time points alerts us to an issue that the post-demanding environment simply captures more dispensing/prescribing of ORS, and thus higher correct care because it captures all actions after the entire visit has been completed. Since providers are as likely to advise or dispense/prescribe ORS before demanding, we believe based on similar pre-demanding rates, if it were a methodological limitation, we would also see similar post-demanding rates across clinics assigned to albendazole versus amoxicillin, but this is not the case. Thus, this may suggest that providers who receive a client demanding a non-efficacious, harmful drug are counteracting that demand by dispensing/prescribing the correct drug. A future study interested in this could conduct experiments on not demanding and demanding as separate visits to ascertain the extent to which this is true. With field constraints, we decided to combine the experiment into the same visit.

[21] See appendix tables 6.A2.5a and 6.A2.5b for regressions with an interaction between being unmarried and AHME treatment.

Poverty

For the experiments conducted for asthma and malaria by SPs portraying poor clients (figures 6.6 and 6.7, respectively), we find that the poor were significantly less likely to receive correct management and significantly less likely to receive any (unnecessary) lab tests for asthma but not for malaria at the α=0.05 significance level. We find the poor were also less likely to receive any unnecessary medicines for both cases.

For the malaria case specifically, SPs portraying poor clients were less likely to receive microscopy than SPs presenting as not poor (coefficient = -0.152, p-value = 0.029; table 6.10, panel a, model 5); however, poor and not-poor SPs did not statistically differ in receiving malaria mRDT (coefficient = -0.023, p-value = 0.724). Although malaria SPs presenting as poor were significantly less likely to receive any unnecessary medicines (figure 6.7), they were as likely to receive AL (coefficient = 0.015, p-value = 0.816; table 6.10b, model 5) when compared to SPs presenting as not poor.

As for differential AHME effects, we find significantly different AHME treatment effects for poor SPs vs. not-poor SPs on our malaria diagnostic test result outcomes (appendix table 6.A2.6b): whether the SP was considered malaria positive unconditional (coefficient = 0.146, p-value = 0.065; model 1) or conditional (coefficient = 0.189, p-value = 0.043; model 2) on receiving a test; whether the malaria RDT was positive (coefficient = 0.439, p-value = 0.007; model 3); and whether the malaria microscopy was positive (coefficient = 0.232, p-value = 0.082; model 4).

The significantly lower levels of correct care provided to the poor emphasizes the importance of social protection and health insurance, particularly for the poor. We would think that NHIF empanelment of the poor would improve affordability of high-quality services, but because we do not have a randomized experiment where the poor case is covered by NHIF, we are not able to causally identify whether this holds true.

6.3 Process Quality Discussion and Conclusion

Based on our framework, structural quality is necessary but not sufficient to improve process quality. In the AHME program, we did not see the significant improvements on some structural quality measures translate into improvements in process quality. In fact, we find that the AHME program significantly reduced correct actions provided to clients, and we also find lower levels of wasteful diagnostic and treatment actions due to the program; however, these results were not significant. We further find large deficits in laboratory quality associated with malaria diagnostics that can have consequences from onward treatment; however, laboratory quality was outside the scope of the AHME intervention. In this section, we explore structural components that could have led to reduced correct and unnecessary care. In particular, we examine the levels of structural quality for provider competence and supplies.

To examine whether provider competence led to lower rates of correct and unnecessary care in the AHME program, we examine provider knowledge for the cases presented by the SPs, as well as time spent working at the clinic and other provider characteristics. Overall, at endline, provider knowledge was high among the AHME control group: 91.9% of providers knew to correctly manage the diarrhea case; 93.4% of providers knew at least one component of correctly managing the family planning case (though only 10.7% knew all four components); and 98.4% knew to correctly manage the malaria case (table 6.12). Across both treatment and control clinics, we find significantly higher rates of knowledge compared to practice among the same set of health care providers (figure 6.8).[22] We find AHME treatment clinics are not statistically different from control clinics at endline concerning knowledge of correct management at the α=0.10 significance level (table 6.13). Higher rates of knowledge than practice are consistent with the know-do gap literature, which also argues that training interventions would not improve outcomes because knowledge is already high (Das et al. 2015; Mohanan et al. 2015).

Further, upon examination of N=880 SP observations where we have data for how long providers have spent working at their respective clinics, we find no correlation between correct management or knowledge of correct management with experience. We do find that providers seen for visits at control clinics had worked at that clinic 0.94 years longer (difference in means p-value = 0.092) than providers seen for visits at treatment clinics. Among AHME control clinics, SPs were seen by providers who had spent an average of 9.57 years (std. dev. 8.25 years, range [0.08, 27]) working at that clinic; among AHME treatment clinics, providers had spent an average of 8.64 years (std. dev. 8.07 years, range [0.25, 29]) working at that clinic.

Given the positive correlation between medical training and knowledge yet little or no correlation between knowledge and experience, Das and Hammer (2014) suggest that younger provider cohorts may be better trained. In this study, in a multivariate regression analysis controlling for SP and case fixed effects, we find that provider knowledge is a significant predictor for correctly managing the cases; however, we do not find provider age, provider gender, and provider qualification to be strong predictors (appendix table 6.A2.10). Thus, similar to Das et al. (2020), we find that formal qualifications are poor predictors of quality.

Since knowledge was high, it is possible that a lack of supplies (stockouts) set up the conditions for lower levels of care in AHME. We do not have evidence supporting the hypothesis that the know-do gap can be explained by stockouts or supply constraints. There are two reasons for this. First, providers can order or prescribe items not in stock or ask clients to return for those items. Our SP data capture whether the provider took these actions and the reasons why, regardless of a stockout.

Second, a closer examination of stockouts demonstrates that between 1%–2% of the clinic sample mentioned stockouts for any of the lab tests or medicines found in the SP data. We cross-examined stockout data from SP reports and clinic administrator self-reports and found low levels of stockouts in the clinic survey for medicines related to the SP experiments. Specifically, AL medicines for malaria were out of stock at 2% of clinics (n=199); paracetamol

[22] See appendix figures 6.A2.1, 6.A2.2, and 6.A2.3 for differences between knowledge and practice for the diarrhea, family planning, and malaria cases, respectively.

was out of stock at 5% of clinics; albendazole was out of stock at 1% of clinics; and amoxicillin was out of stock at 6% of clinics. The findings from our evaluation sample, which covers 35 of Kenya's 47 counties, are inconsistent with research conducted across multiple low- and middle-income countries that have found drug availability to be as large a barrier as provider availability (Wilson 2020).[23]

Despite some significant improvements to structural quality, we do not find strong evidence that AHME resulted in higher levels of process quality outcomes for outpatient services. In our study context, SPs excelled at revealing what occurs when the same patient visits a random subset of the countrywide clinic sample, presenting the same condition as a new walk-in patient. By leveraging the randomized field experiment and experimentally varying SP case presentation, we were able to identify the effects of the program and certain client characteristics (age, gender, demanding medicines, marital status, and poverty) on process quality measures. Strikingly, we find that regardless of the AHME program, the poor receive lower levels of correct and unnecessary care, and there were large deficits in laboratory quality. However, what this part of the study cannot show or reflect is the patient mix and patient sorting that occurs in reality. In the next chapter, we examine health care outcomes related to quality of care with the combination of several data sources.

[23]As mentioned in a footnote in Chapter 4, we do not believe provider availability to be a large issue either, since 1174 (98%) of 1195 of all visits were successfully conducted on the 1st attempt.

Chapter 7

Health Care Outcomes

This chapter focuses on the third type of quality of care in our framework: health care outcomes. Specifically, we seek to answer the following two quality of care research questions related to health care outcomes:

1. What was the effect of the AHME program on client perceptions of amenities and client satisfaction?
2. What was the effect of AHME on the perception of clinic quality among households?

From our analyses, we find few effects of AHME on client perceptions of amenities and client satisfaction. We do not find evidence for AHME effects on client experience in terms of number of clients in the waiting room (a proxy for wait time), time spent with the provider, and total time spent at the clinic. We also do not find evidence for households perceiving any difference in quality between AHME treatment and control clinics. The remainder of the chapter details quality of care related to health care outcomes measurement and analysis, results, and discussion.

7.1 Measurement and Analysis of Health Care Outcomes

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

Data Sources

To understand health care outcomes, we use AHME evaluation endline data from SPs, exit interviews with actual clients, and the household survey. To understand client perception of clinic amenities and client satisfaction, we use data from SPs and exit interviews with actual clients. Section 6.1 contains details on the SP surveys. For the client exit interviews, clients were recruited at endline from AHME treatment and control clinics and interviewed to understand client perceptions of clinic amenities, using the same questions administered to the SPs. For the household survey, clients were recruited at baseline from AHME treatment and control clinics, and interviewed by the field team to understand demographics, health services utilization, and well-being of client households. At endline, we asked one adult respondent per household his or her perceptions of health clinic quality. Gertler et al. (2020) contains additional details for the endline household survey.

Health Care Outcome Measures

This section first describes the health care measures from the SP and exit interview data, followed by the health care measures from the household survey.

Measures from SP and Exit Interview Surveys

Components of process quality, such as clinic amenities, play a role in health care outcomes. For example, clinic amenities influence client perceptions and are known to be positively correlated with client satisfaction (Goldman, Vaiana, and Romley 2010). We consider both client perceptions and client satisfaction to be health care outcomes, based on the quality of care framework in figure 3.1. To examine the impact of the AHME package and in particular SafeCare's targeting of client amenities on client perceptions and client satisfaction, we assessed the following 9 client perceptions of clinic amenities from the SPs' perspective: whether (1) the clinic was clean; (2) waiting time was appropriate; (3) providers were courteous and respectful; (4) SPs had enough privacy; (5) providers spent sufficient time; (6) operating hours of the clinic were appropriate; (7) SPs completely trusted the provider's treatment decision; (8) registration fees were reasonable; and (9) medicine and drug fees were reasonable. We coded these perceptions as 0-1 binary measures. Measuring and classifying these components is challenging because many of the features can be viewed as objective (such as when SPs are trained to understand what "clean" is), though they are inherently subjective. We refer to these 9 measures as health care outcomes, though some could qualify as health care structures, processes, or outcomes.

Because SP satisfaction cannot be interpreted as client satisfaction, we also report a tenth binary measure, the SP's assessment of whether the provider did a good job explaining treatment and care options. Because SPs are not actual clients, we assessed all 9 binary perception measures related to amenities and the tenth measure of whether the provider did a good job explaining treatment and care options with data from client exit interviews. These 10 binary measures are drawn and adapted from surveys from the World Bank Service Delivery Indicators initiative (Evans et al. 2020).

In addition, we examined the effect of AHME on SP satisfaction, based on SPs reporting their satisfaction on a 1 to 10 scale, with 1 being least satisfied and 10 being most satisfied.

We transformed the SP-reported measures to z-scores with AHME control group mean zero.

Measures from the Household Survey

To further understand actual client perceptions of clinic quality, we used data from the household survey. Respondents were asked to rate and rank how the treatment or control clinic from which their household was recruited at baseline (that is, the index clinic) compared to other health clinics they had recently visited. The number of comparison clinics varied, ranging from 0 to 4 for all households. We asked household respondents for their perception of clinic quality for three hypothetical visits: a child curative service visit, a child preventative service visit, and an adult curative service visit. For each of these services, respondents first rated the quality of care at the index clinic and each comparison clinic from 1 (lowest quality care) to 10 (highest quality care). Then they were asked to rank the clinics from highest quality to lowest quality. Because the numbers of clinics in clients' comparison groups varied, we created an indicator variable for whether the index clinic was in the highest half of their ranking. Both the rating and ranking outcomes were transformed to z-scores with a mean of zero for the control group (mean of control group zero). Clients were asked to rate and rank each of three hypothetical visits. We created pooled models by combining the responses for each client and added indicators for the type of hypothetical visit in our regression models.

Estimating Equations

For analyses at the SP level, we used the linear regression model described in section 6.1. For analyses at the client visit and household individual level, we used the following conditional linear regression (equation 7.1), including additional covariates that could impact the outcome of interest:

$$Y_i = \beta_0 + \beta_1 AHME\ Treatment_i + \beta_2 X_i + \epsilon_i, \tag{7.1}$$

where β_0 is the intercept, and $AHME\ Treatment_i$ is a binary treatment indicator for whether the clinic associated with the individual received AHME ($AHME\ Treatment_i = 1$) or was assigned to the control arm ($AHME\ Treatment_i = 0$). The coefficient on the AHME Treatment indicator variable, β_1, is interpreted as the impact of the AHME intervention on the outcome of interest (Y_i), and ϵ_i is a normally distributed error term for observation i, clustered at the index clinic level. Covariate X_i is added to control for potential baseline imbalances, and β_1 is interpreted as the impact of $AHME\ Treatment_i$, conditional on X_i.

7.2 Health Care Outcome Results

This section reports the effects of AHME on client perceptions of clinic amenities and client satisfaction from both SP and client exit interviews, followed by household perceptions of clinic quality.

Table 7.1 shows the effects of AHME treatment on client perceptions of clinic amenities from the perspective of SPs. First, we do not find evidence supporting the hypothesis that the AHME program improved amenities, based on our 9-item index (coefficient = 0.008, p-value = 0.471; model 1). We find that AHME program assignment significantly improved privacy by 6.6% (coefficient = 0.054, p-value = 0.001; model 5) from the AHME control group mean of 0.822. There were no significant effects of AHME on whether the clinic was clean, the wait time was appropriate, providers were courteous and respectful, providers spent sufficient time, SPs completely trusted the provider's treatment decision, or registration fees or medicine fees were reasonable (models 2–4, and 6–10). When we included interactions between AHME treatment and each SP experiment, we find that SPs perceived AHME treatment clinics to have more privacy and more adequate operating hours at the α=0.01 and α=0.10 significance levels, respectively, when compared to AHME control clinics (appendix table 7.A1.1, models 4 and 6). Since SPs are not actual clients, we also examined the same outcomes among actual clients in exit interviews. We did not find any program effects for the same 9-item amenities index and for each component comprising the index at the α=0.10 significance level (table 7.2).

Next, we examined client satisfaction and whether providers did a good job explaining treatment and care options from the perspectives of both SPs and actual patients.[1] We do not observe any AHME program effects for SP satisfaction for any of the SP cases or the pooled SP data at the α=0.10 significance level (table 7.3). For whether providers did a good job explaining treatment and care options, SPs presenting the family planning case rated AHME treatment clinics statistically higher (coefficient = 0.150, p-value = 0.034) than AHME control clinics (table 7.4, model 2). However, for the other cases and pooled SP data, we do not find statistically significant differences between AHME treatment and control clinics at the α=0.10 level (table 7.4, models 1, 3, 4, 5). When we analyzed exit interview data, there was no evidence that the AHME program had an effect on clients believing that providers did a good job explaining treatment and care options at the α=0.10 significance level (table 7.4, model 6).

Further, we find no effect of AHME on household perceptions of clinic quality. Clients of AHME clinics at baseline did not rate those clinics higher than comparison clinics (p-value = 0.471; table 7.5, model 1). Clients in the AHME treatment group were also not more likely to rank the index clinic as high quality than clients in the control group (p-value = 0.302; table 7.5, model 2).

7.3 Health Care Outcomes Discussion and Conclusion

We leveraged SP, exit interview, and household survey data from the AHME evaluation endline activities to assess health care outcomes. First, we find that the AHME program

[1]Appendix tables 7.A1.2a and 7.A1.2b show models including interactions between AHME treatment and each SP experiment for SP satisfaction and whether the provider did a good job explaining treatment and care options, respectively.

improved client perceptions of clinic privacy in the SP data, but not in the exit interview data. We do not find any program effects on client perceptions of other clinic amenities or client satisfaction. We find that SPs portraying the family planning scenario believed providers at AHME treatment clinics did a better job explaining, but we did not detect any effects for this outcome among SPs portraying other cases or among actual clients. At the household level, we do not find evidence that household members observed any differences in clinic quality between AHME treatment and control clinics.

To assess whether our findings of no program effects on client perceptions were biased, we examine process quality measures often assessed in the literature and related them to the client's experience with SP and exit interview data (Goldman, Vaiana, and Romley 2010). For whether waiting time was appropriate, we assessed the effects of AHME treatment on the number of clients waiting at the clinic (SP and exit interview data) and minutes spent waiting (exit interview data). For whether providers spent sufficient time, we assessed minutes spent with the provider (SP data). We also assessed the total time spent at the clinic (SP data). We transformed these SP- and client-reported measures to z-scores with AHME control group mean zero. When we compare AHME treatment and control clinics, we do not find that the AHME program changed wait times or the number of people waiting at the clinic, the amount of time spent with the provider, or the amount of time spent at the clinic, including wait time (table 7.6).[2]

Lastly, we do not find evidence that household perceptions of clinic quality were different for AHME treatment and control clinics. Together with findings from chapter 6 that showed that the AHME program had a negative effect on correct actions, we observe that clients do not recognize this change in the market. This demonstrates a market failure, which we examine and discuss in the final chapters.

[2]See appendix tables 7.A1.3a, 7.A1.4a, and 7.A1.5a for case-wise models without and appendix tables 7.A1.3b, 7.A1.4b, and 7.A1.5b with AHME treatment and SP experiment interactions.

Chapter 8

Provider Preferences: Altruism vs. Self-interest

Growth in the private health sectors in low- and middle-income countries has been previously attributed to rising incomes and the inability of the public sector to meet expectations (Mills 2014). In previous chapters, we find that clients and households did not recognize the observed change due to AHME in the private sector for process quality. We conjecture that this market failure could be due to medical information asymmetry 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 patients.

This chapter describes our analysis of providers' social preferences-altruism and self-interest in particular. Recognizing that preferences and attitudes shape behaviors, we begin this chapter with SP narratives during fieldwork that identify altruistic and profit-driven behaviors among providers in the AHME evaluation sample. These narratives motivate our interpretation of self-interest as "profit driven" in the context of the Kenyan private sector. Then, we turn to our analysis on provider preferences towards altruism and self-interest (for-profit) elicited from a modified dictator game matched to SP visits. In so doing, we answer the question, *Do the effects of AHME on correct case management and how much a patient is charged vary by a provider's profit-driven instincts?*

8.1 Narratives from Fieldwork

First, we examined post-SP visit narratives from SP pilot and fieldwork. We find anecdotal evidence for both altruistic and profit-driven behaviors. Regarding the former, we note that a small number of SPs portraying the poor case variant reported that concerned providers paid for necessary lab tests and medicines. Conversely, a small number of SPs reported instances in which the electricity did not work while the malaria microscopy test was being processed, rendering the ability to use a microscope impossible-although providers then proceeded to deliver a (false) positive test result to the SP.[1]

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

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

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

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

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

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

8.2 Altruism vs. Self-Interest: Results from a Modified Dictator Game

Next, we explore the effects of AHME across levels of social preferences among providers with data from a modified dictator game. Following literature that suggest provider preferences play a role in decision making among health care providers, we examine provider preferences, specifically distributive social preferences. Distributive social preferences-including altruism and spitefulness, fairness and aversion to inequity, and concerns regarding efficiency-relate to an individual's preferences toward the distribution and magnitude of payoffs among others on a range of issues (e.g., social security, healthcare, benefits). In the context of the provision of quality health care in the private sector, we focus on altruistic social preferences - a type of distributive preference that we hypothesize to govern the trade-offs that a health care provider makes between his payoffs and the payoffs to his clients (Fisman, Jakiela, and Kariv 2017). To elicit altruism, we use a real-stakes modified dictator game, and in our case, the dictator is a provider at one of the clinics in the AHME evaluation sample.

Dictator games provide a nonstrategic environment to elicit social preferences. We adapt a modified dictator game originally modified from Andreoni and Miller (2002) in several subsequent studies (Li, Dow, and Kariv 2017; Fisman, Jakiela, and Kariv 2017; Balakrishnan, Haushofer, and Jakiela 2020; Jakiela 2013). In our game, the provider, who is given 5 different scenarios of real or fictitious clients with different characteristics (e.g., poor), determines how much of an endowment he wants to give to the client and how much he wants to keep for himself.

Our modified dictator game consists of decisions that are a zero-sum game and decisions that are a positive-sum game–in other words, 100 Kenyan Shillings (KSH) lost for the dictator is 100 KSH gained by the recipient (zero sum) versus 100 KSH lost for the dictator increases the recipient's payoff to $200, 300, 400, 500$ KSH (positive sum). In other words, depending on the scenario, we apply a 1x, 2x, 3x, 4x, or 5x multiplier. In the literature, the larger the multiplier, the more the dictator is likely to give. In our game, each dictator (provider) divides the constant endowment w of 1000 KSH between themselves and a recipient. This can be expressed as

$$\pi_s \ + \ p_o\pi_o \ = \ w \tag{8.1}$$

where π_s is the payoff to self (provider), π_o is the payoff to other (a real or fictitious patient), and p_o takes on values $1, 2, 3, 4$, or 5. When p_o does not equal 1, then the provider faces a trade-off between equity and efficiency. Over the course of the game, the provider makes a total of 19 decisions with each provider making the same 19 decisions. The 19 decisions are spread across 5 different case scenarios: 5 decisions (p_o takes on values 1, 2, 3, 4, 5) for a real client, 3 decisions (p_o takes on values 1, 3, 5) for a fictitious childhood diarrhea client, 3 decisions (p_o takes on values 1, 3, 5) for a fictitious family planning client, 3 decisions (p_o takes on values 1, 3, 5) for a fictitious malaria client, and 5 decisions (p_o takes on values 1, 2, 3, 4, 5) for a fictitious malaria client who is also poor. 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. At the end of the game, a lottery determined which scenario would play out, giving the respondent and a real client an opportunity to win mobile money corresponding to the decision made by the respondent (Andreoni and Miller 2002; Fisman, Jakiela, and Kariv 2017; Balakrishnan, Haushofer, and Jakiela 2020; Li, Dow, and Kariv 2017; Jakiela 2013). Appendix 8.A1 contains further details of the protocol for the modified dictator game, descriptive tables and figures summarizing the data, consistency checks, and supplemental tables and figures from our analysis. Appendix table 8.A1.1 summarizes the game's scenarios and decisions.

Following Jakiela (2013), we first report summary statistics on individual decisions. Our first outcome of interest is budget share spent on tokens for other, defined as $p_o\pi_o/(\ \pi_s\ + p_o\pi_o)$. In Table 8.1, we see that the average client budget share was 54% for all 19 decisions made by each respondent of the provider survey (N=302).[2]

Using the provider survey respondents, Table 8.2 shows AHME's effects on client budget share in the modified dictator game. Controlling for features of our modified dictator game (e.g., multipliers, client scenarios, whether the scenario was described as poor), we find that AHME reduced client budget share by 16.0% (coefficient=-0.090, p-value<0.001) compared to the AHME control group mean of 0.563. When controlling for whether the respondent was the clinic in-charge, we find that AHME reduced client budget share by 12.8% (coefficient=-0.072, p-value<0.001).

Next, we turn to estimating our social preference parameter of interest: altruism. Following the literature, to estimate our parameter, we perform checks on our data to confirm each respondent's 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, "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 utility function, " meaning the utility function is piecewise, linear, continuous, increasing, and concave (Li, Dow, and Kariv 2017).

In order to check for consistency, we assess how the data comply with GARP by calculating Afriat's critical cost efficiency index (CCEI) (Fisman, Jakiela, and Kariv 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. Following literature that uses the threshold of 0.85 for Afriat's CCEI (Fisman, Jakiela, and Kariv 2017), we find that our data comply with GARP (see "Afriat CCEI Violation" in Table 8.1). In fact, all of our data comply with GARP at the 0.95 threshold level, and 52 (17.3%) of 300 provider survey respondents (corresponding to 142 (16.7%) of

[2] In table 8.1, we provide summary statistics from respondents of the provider survey (panel a), and from the respondents of the provider survey matched to providers seen in SP visits, where the same interviewed provider could have successfully seen and been identified for one or more SP visits (panel b).

the 850 SP visits that match to the provider survey) violate GARP at a CCEI threshold of 1.00. We report the proportion of observations that are related to GARP violations at Afriat's CCEI thresholds of 1.00, 0.95, and 0.80 in Table 8.1, panels a and b.[3,4] Because we do not have data that violates Afriat's threshold, we use all our data, and we examine three social preference parameters estimated following Andreoni and Miller (2002) and Li, Dow, and Kariv (2017) (see Appendix 8.A1 for estimation details): (1) $\alpha \in [0, 1]$, which denotes the range from altruistic to fair-minded to self-interested, (2) $\rho \leq 1$, which denotes tendencies towards efficiency versus equality, and (3) $\sigma = 1/(\rho - 1) \leq 0$, which denotes the (constant) elasticity of substitution. We do not examine ρ or σ because of the lack of variation in these parameters. Instead, we focus on social preference parameter α and its ability to discern providers along the spectrum of profit-oriented to altruistic preferences.[5]

We want to test the hypothesis that the negative effects of AHME on correct case management can be explained by providers with preferences towards self-interest. Given that providers partake in health care provision in the private sector and based on narratives from fieldwork, we interpret self-interest in this setting as "profit-driven preferences." To test our hypothesis, we examine whether there was heterogeneity in AHME's effects across providers who were profit-driven (self-interested) versus those who were more altruistic. We find evidence that supports our hypothesis.

Following the literature, 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 plot a kernel density of the social preference parameter alpha conditional on AHME treatment status (figure 8.1). We observe that the distributions for alpha are different between [0.4, 1], which corresponds to the range of providers who are fair-minded to self-interested. Interestingly, we see that that providers at AHME control groups are more heavily concentrated around fair-mindedness, and providers at AHME treatment groups are more heavily concentrated near perfect self-interest.

Since there is no determined cutoff in the literature for identifying the providers who are the most self-interested, we take the following approach: we sequentially identify the bottom 50th percentile, bottom 25th percentile, and then the bottom 20th percentile of providers on our altruism parameter elicited from the modified dictator game (e.g., the bottom 20th percentile is the 20% least altruistic, equivalent to the 20% most self-interested). Using a multivariate OLS regression model (similar to Equation 6.1 in Chapter 6) with data at the SP-provider level, we show the effect of AHME on correct case management

[3]Observations are not consistent within panels, because we are only able to determine Afriat's CCEI threshold violations and social preference parameters for provider survey respondents who conducted all decisions in the modified dictator game.

[4]See appendix table 8.A1.2 for descriptive statistics on provider allocations and client budget shares across decisions from the modified dictator game.

[5]Appendix figure 8.A1.1, panels a, b, and c show histograms of social preference parameters alpha, rho, and sigma, respectively, for provider survey respondents.

without controlling for altruism (Table 8.3, model 1), and then to the same model, we include indicators for whether the SP saw a provider in the 50% least altruistic group (Table 8.3, model 2), 25% least altruistic group (Table 8.3, model 3), and 20% least altruistic group (Table 8.3, model 4). We note that the magnitude of the AHME treatment coefficient gets smaller and no longer is significant at the 5% or 10% significance level as we narrow in on the most self-interested providers.

To test our original hypothesis of whether the most profit-driven providers reduce correct care due to AHME, we examine the heterogeneity of treatment effects on correct case management by including an interaction term between AHME treatment and each indicator representing whether the SP saw a provider falling into the 50% least altruistic, 25% least altruistic or 20% least altruistic group. As we narrow in on the least altruistic providers with our approach, we find that the most profit-driven providers at AHME clinics were the ones to significantly reduce correct care (see Table 8.4).[6]

Adjusting for this in our model demonstrates that the providers who are among the 80%, 75% and even 50% most altruistic do not contribute to the significant reduction in correct care due to the AHME program, since the coefficient on the AHME treatment indicator adjusting for the interactions is no longer significant at the 10% significance level (Table 8.4, models 3, 4, 5). In parallel, what is striking is that the magnitude of the coefficients on the interactions get larger and increasingly significant (coefficient on AHME treatment and least altruistic (50%) in Model 3: -0.159, p-value = 0.074, coefficient on AHME treatment and least altruistic (25%) in Model 4: -0.189, p-value = 0.073, coefficient on AHME treatment and least altruistic (20%) in Model 5: -0.227, p-value = 0.063, respectively; Table 8.4, models 3, 4, 5).

Table 8.5 displays the heterogeneity of AHME treatment effects on total prices paid (arcsinh transformation) with interactions between AHME treatment and whether providers are among the 50%, 25%, or 20% least altruistic as described for our analysis on correct case management. We see that using the model from Chapter 6 Equation 6.1 on the outcome of total price paid by SP (in Kenyan Shillings, KSH) transformed to hyperbolic arcsine following Bellemare and Wichman (2020), we do not find evidence that the AHME program changed prices for patients at the 10% significance level (Table 8.5, model 1). However, we find that SPs seeing the 20% least altruistic providers at AHME treatment clinics paid 121.2% more than the 20% least altruistic providers at AHME control clinics, and this was significant at the 1% significance level (semi-elasticity of least altruistic (20%) in AHME:

[6]More specifically, in Table 8.4, model 1, we first include providers correctly identified and replacements for those providers who were not identified (matched on clinic and whether that provider sees patients for that service). Models 2-5 restrict the sample to SP visits conducted by correctly identified and matched providers from the provider survey. Model 1 does not include altruism parameters. Model 2 does not include altruism parameters and restricts the sample as described. Models 3, 4, and 5 include indicators for whether the provider seen by the SP is among the 50%, 25%, and 20% least altruistic, respectively of our continuous altruism parameter, and each model includes an interaction between AHME treatment (0 if clinic was in AHME control and 1 if clinic was in AHME treatment) and the indicator for the whether the provider fall in the 50%, 25%, and 20% least altruistic group in order to identify heterogeneous treatment effects for the most for-profit providers versus those who are not.

1.212, p-value = 0.003; Table 8.5, model 5). For this same model, we see that AHME treatment clinics increased prices by 2% compared to the AHME control group mean, but this was not significant at the 10% significance level (semi-elasticity of AHME: 0.020, p-value = 0.693; Table 8.5, model 5).

Our findings suggest that distributive preferences, particularly altruism and self-interest, among private providers strongly play a significant role in the provision of quality and affordable care. We find that AHME possibly changed the incentives of providers (i.e., made them more profit-driven) without changing financial incentives, as supported by figure 12.1. However, it could be possible that providers varied their preferences before participating in the AHME program and those that are more profit-driven responded to the program in a different way than other providers in the program. Both interpretations imply that business management interventions should be better aligned with process quality measures. The latter could alternatively imply that interventions should either be better tailored to provider preferences or conduct better signaling or screening measures to select providers according to their preferences. Understanding provider altruism can thus help identify strategies to better align future interventions with process quality measures and to strengthen data captured for monitoring purposes.

Chapter 9

Discussion

This chapter discusses our findings, compares the levels of quality we found to standardized patient (SP) studies in other settings in Africa and India, addresses the limitations of our quality measures, and discusses policy implications for scale-up, program design, and research design.

9.1 Comparisons to Other Studies and Settings

We first assess how the levels of care we find compare to other settings. We were able to compare our diarrhea and asthma case results to other SP studies conducted in Nairobi and India. First, Daniels et al. (2017) conducted an SP study in Nairobi and found that 73% of childhood diarrhea cases (62% in public clinics; 82% in private clinics) and 82% of asthma cases (79% in public clinics; 82% in private clinics) were correctly managed by clinics that were purposively sampled (selected based on characteristics of a population and the purpose of the study). Their results are higher than what we find in the AHME evaluation clinics for diarrhea (69%) and asthma (44%). We attribute these differences to the study sampling frames and recruitment procedures; however, future analyses can examine this further.

At the country level, Kenya has similar or higher rates of knowledge and practice than reported in studies of rural and urban India. Das et al. (2012) find that 12% of childhood dysentery and 48% of asthma SP cases were managed correctly. Mohanan et al. (2015) find that only 17% of providers actually offered oral rehydration salts (ORS) to the childhood diarrhea SP case, despite 72% of the same providers reporting that they would. For ORS in our study, providers knew to give ORS for 81% of diarrhea SP interactions, with 38% of those interactions actually resulting in offering ORS. As we write this, there are ongoing studies that will produce comparable SP data in sub-Saharan settings with similar case scenarios, but until they are published, we have little data to compare our quality of care findings across Sub-Saharan Africa. However, the World Bank's Service Delivery Indicators survey offers provider competence measures using direct observation and vignettes (World Bank Group, 2017 (accessed April 21, 2020)). The World Bank results demonstrate that provider competence in Kenya is higher compared to provider competence in Tanzania or Senegal for five conditions: while diagnostic accuracy was 72.2% in Kenya, with 43.7% adherence to clinical guidelines, it was 57% for Tanzania, with 35% adherence; and 34% in Senegal, with 22% adherence (Bold et al. 2011; Gayle and Pimhidzai 2013; Das and Hammer 2014).

9.2 Limitations of Our Quality Measures

The SP method and clinical vignettes are considered the gold standard methods for assessing provider practice and provider knowledge; however, both methods have limitations. Although the methods provide a "standardized" client scenario to evaluate multiple providers at multiple locations, they do not represent the heterogeneity of client mix and client sorting that occurs in an entire health care market. Since we examine private primary health care clinics in our study, we argue that the minimal, essential actions we have identified to correctly manage the childhood diarrhea, family planning, mild asthma, and malaria scenarios are indeed minimal, basic essential actions that outpatient clients should receive in accessing high-quality primary care in any setting.

Further, cited in the literature, the limitations of the SP method—that it is limited to one-time, walk-in visits for services that do not result in invasive procedures; is related to tracer health conditions that have been validated for ethical research; and cannot be related to ailments that require established client services or follow-up visits—proved to be favorable to assess the quality of walk-in outpatient services of interest in the AHME program, and the SP method was demonstrated to be a rigorous way to capture the outcomes of interest (Kwan et al. 2019). It is concerning that providers knew how to correctly manage the scenarios we presented, but did not act on that knowledge and manage cases correctly in practice.

Regarding our study of client perceptions of clinic amenities, one important caveat is that SP satisfaction should not be interpreted as actual patient satisfaction. The SP's subjective opinion is influenced by several factors that differ from actual patients, such as the training they have undergone to be actors and whether the SP knows the provider is clinically doing a good job. This bias is particularly salient for the case of malaria, where any SPs receiving a positive test are more inclined to rate the clinic and the provider lower than when visits resulted in a negative test result. For this reason, we capture client/patient perspectives of client amenities from exit interviews with real clients/patients, as well as households in index clinic catchment areas.

With respect to the potential social desirability bias of self-reported knowledge surveys and the Hawthorne effect associated with administrative data, we conclude that SP data, complemented with other quality of care measures, were a powerful multi-faceted approach to capture both improvements and deficiencies in care.

9.3 Implications for Scale-Up, Program Design, and Research Design

Although the AHME program improved dimensions of clinic management and structural quality structural quality measures, we do not find that these structural improvements translated into better process quality or health care outcomes for outpatient services. Despite high levels of provider knowledge in the market and substantial improvement in AHME for structures to support patient rights and record keeping, AHME treatment clinics provided on average significantly worse correct case management and slightly lower, yet not significantly different, levels of wasteful services compared to control clinics. However, reductions in correct care were made by the least altruistic providers, who also charged more for services. We conclude that the considerable amount of effort in this intervention purchased many improvements, but allowed for the most profit-driven providers to reduce services, including minimally essential actions for outpatient, and these reductions were met with significantly higher prices. We do not know whether this holds for inpatient services, but only a minority of clinics reported providing inpatient services in the clinic survey: 28 of 108 of the treatment clinics, and 22 of 89 control clinics. Understanding provider altruism can thus help identify strategies to better align future private sector engagement efforts with process quality measures and to strengthen data capture for monitoring purposes.

Not only did we observe reductions in minimal and essential actions related to correct care management in the context of the AHME program, but we also found striking deficits in laboratory quality and in the care provided to the poor. We attribute these lower levels of quality to (1) the intervention's lack of focus on process quality, specifically related to the content of care and laboratory quality for the health services of interest in AHME, and (2) the difficulties for program implementers to measure these components with methods that are easy to implement, yet rigorous and unbiased. That we find the poor get worse care in general emphasizes the importance of social protection programs. However, we also find that the poor get worse care in a program (AHME) that specifically aims to improve private sector access to quality of care for the poor—though we cannot precisely examine the care received by a poor individual covered by NHIF.

We observe that an intervention on training providers to improve knowledge would not be successful for improving correct diagnostic and treatment for clients or improving laboratory quality, in the case when high levels of knowledge exist but practice of correct actions is low (and can be further reduced). Given our findings, we caution against the combined scale-up of SafeCare quality improvement interventions, Social Franchising, Business Support, and NHIF Empanelment Support interventions for improving quality of care for clients without more rigorous attention to private provider preferences *and* process quality measures — beginning with the content of clinical care and laboratory quality.

Chapter 10

Conclusion

By taking advantage of a randomized experimental design, we were able to assess the effects of AHME on various types of quality of care: structures, processes, and health care outcomes. Based on Donabedian's research on quality of care, these three types of quality of care proved to be a comprehensive framework to apply in answering the question, *Can we improve quality of care in private health sectors?* In the context of a quality improvement program in the private sector of Kenya, we find that in some ways we can improve quality of care, but heterogeneity in provider and patient characteristics can influence our ability to improve correct care, and in some situations, these characteristics can lead to reductions in care quality at higher prices to patients. Because of the complexity of improving content of care and laboratory quality, we advise on interventions that heavily emphasize management and accountability interventions for content of care and laboratory quality with effective regulatory and monitoring components. Quality improvement programs may also want to invest in identifying and/or changing norms, such as towards fair-mindedness or altruistic preferences. We also advise against market-based interventions and basic performance incentives monitored with observation audits conducted by program staff.

The research designs for both the AHME impact evaluation surveys, together with the data from the clinic, SP, client exit, and household surveys were powerful for assessing the impact of the AHME program on health care structures, processes, and outcomes with high internal validity. The AHME impact evaluation is the first study that examines truly comparable data among a clinic sample on structural and process quality using a methodologically powerful combination of provider vignettes, the gold standard SPs, real client exit interviews, and a household survey across the country of Kenya. This study is additionally the first study to identify the causal impact of various client characteristics on quality care.

Chapter 11

Research Team

This research was led by Ada Kwan under the supervision of Paul J. Gertler. Members of the AHME impact evaluation research team include: Claire Boone, David Contreras-Loya, Rita Cuckovich, Joshua Gruber, and Nicole Perales. Fieldwork was coordinated by Ada Kwan and supervised by Andrew Muriithi, Pheliciah Mwachofi, Purity Kimuru, Rodgers Kegode, and Salome Omondi.

Bibliography

AHME Program. 2016. *AHME Business Improvement Plan Example.*

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

Arrow, Kenneth J. 1963. "Uncertainty and the welfare economics of medical care." *The American economic review* 53 (5): 941–973.

———. 1998. "What has economics to say about racial discrimination?" *Journal of economic perspectives* 12 (2): 91–100.

Balakrishnan, Uttara, Johannes Haushofer, and Pamela Jakiela. 2020. "How soon is now? Evidence of present bias from convex time budget experiments." *Experimental Economics* 23 (2): 294–321.

Becker, Gary S. 2010. *The economics of discrimination.* University of Chicago press.

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

Benjamini, Yoav, and Yosef Hochberg. 1995. "Controlling the false discovery rate: a practical and powerful approach to multiple testing." *Journal of the Royal statistical society: series B (Methodological)* 57 (1): 289–300.

Bertrand, Marianne, Dolly Chugh, and Sendhil Mullainathan. 2005. "Implicit discrimination." *American Economic Review* 95 (2): 94–98.

Bertrand, Marianne, and Esther Duflo. 2017. "Field experiments on discrimination." In *Handbook of Economic Field Experiments,* 1:309–393. Elsevier.

Birnbaum, A Lord. 1968. "Some latent trait models and their use in inferring an examinee's ability." *Statistical theories of mental test scores.*

Bold, Tessa, Jakob Svensson, Bernard Gauthier, Ottar Maestad, and Waly Wane. 2011. "Service delivery indicators: pilot in Education and Health care in Africa." *CMI Report.*

Christian, Carmen, Ulf-G Gerdtham, Dumisani Hompashe, Anja Smith, and Ronelle Burger. 2018. "Measuring quality gaps in TB screening in South Africa using standardised patient analysis." *International journal of environmental research and public health* 15 (4): 729.

Currie, Janet, Wanchuan Lin, and Juanjuan Meng. 2014. "Addressing antibiotic abuse in China: An experimental audit study." *Journal of development economics* 110:39–51.

Currie, Janet, Wanchuan Lin, and Wei Zhang. 2011. "Patient knowledge and antibiotic abuse: Evidence from an audit study in China." *Journal of health economics* 30 (5): 933–949.

Daniels, Benjamin, Amy Dolinger, Guadalupe Bedoya, Khama Rogo, Ana Goicoechea, Jorge Coarasa, Francis Wafula, Njeri Mwaura, Redemptar Kimeu, and Jishnu Das. 2017. "Use of standardised patients to assess quality of healthcare in Nairobi, Kenya: a pilot, cross-sectional study with international comparisons." *BMJ global health* 2 (2): e000333.

Daniels, Benjamin, Ada Kwan, Srinath Satyanarayana, Ramnath Subbaraman, Ranendra K Das, Veena Das, Jishnu Das, and Madhukar Pai. 2019. "Use of standardised patients to assess gender differences in quality of tuberculosis care in urban India: a two-city, cross-sectional study." *The Lancet Global Health* 7 (5): e633–e643.

Das, Jishnu, Benjamin Daniels, Monisha Ashok, Eun-Young Shim, and Karthik Muralidharan. 2020. "Two Indias: The structure of primary health care markets in rural Indian villages with implications for policy." *Social Science & Medicine:* 112799.

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

Das, Jishnu, Jeffrey Hammer, and Kenneth Leonard. 2008. "The quality of medical advice in low-income countries." *Journal of Economic Perspectives* 22 (2): 93–114.

Das, Jishnu, Alaka Holla, Veena Das, Manoj Mohanan, Diana Tabak, and Brian Chan. 2012. "In urban and rural India, a standardized patient study showed low levels of provider training and huge quality gaps." *Health affairs* 31 (12): 2774–2784.

Das, Jishnu, Alaka Holla, Aakash Mohpal, and Karthik Muralidharan. 2016. "Quality and Accountability in Health Care delivery: audit-study evidence from primary care in India." *American Economic Review* 106 (12): 3765–99.

Das, Jishnu, Ada Kwan, Benjamin Daniels, Srinath Satyanarayana, Ramnath Subbaraman, Sofi Bergkvist, Ranendra K Das, Veena Das, and Madhukar Pai. 2015. "Use of standardised patients to assess quality of tuberculosis care: a pilot, cross-sectional study." *The Lancet infectious diseases* 15 (11): 1305–1313.

Das, Jishnu, Liana Woskie, Ruma Rajbhandari, Kamran Abbasi, and Ashish Jha. 2018. "Rethinking assumptions about delivery of healthcare: implications for universal health coverage." *BMJ* 361:k1716.

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

———. 1978. "The quality of medical care." *Science* 200 (4344): 856–864.

———. 1985. "The methods and findings of quality assessment and monitoring: an illustrated analysis." *Journal for Healthcare Quality* 7 (3): 15.

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

Evans, D., F. A. Dunsch, E. Eze-Ajoku, M. Macis, and Wang Q. 2020. "Service Delivery Indicators: More than Just Results." Presented at Measuring the Quality of Social Service Delivery Symposium, Washington, DC.

Evans, Timothy, Margaret Whitehead, and Finn Diderichsen. 2001. *Challenging inequities in health: from ethics to action.* Oxford University Press.

Fisman, Raymond, Pamela Jakiela, and Shachar Kariv. 2017. "Distributional preferences and political behavior." *Journal of Public Economics* 155:1–10.

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

Gayle, M, and O Pimhidzai. 2013. *Education and health services in Kenya: Data for results and accountability. Service delivery indicators.*

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

Gertler, Paul, Claire Boone, David Contreras-Loya, Rita Cuckovich, Joshua Gruber, Ada Kwan, and Nicole Perales. 2020. *African Health Markets for Equity Quantitative Evaluation Final Report, 2013 – 2019.*

Goldman, Dana P, Mary Vaiana, and John A Romley. 2010. "The emerging importance of patient amenities in hospital care." *The New England journal of medicine* 363 (23): 2185.

Goodman, Catherine, Christina Makungu, Timothy Powell-Jackson, Jessica King, and Nicole Spieker. 2017. "A study to evaluate the effectiveness of SafeCare in Tanzania: Protocols for the use of Standardized Patients."

Hurley, Jeremiah. 2000. "An overview of the normative economics of the health sector." In *Handbook of health economics,* 1:55–118. Elsevier.

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

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

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

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

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

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

King, Jessica JC, Jishnu Das, Ada Kwan, Benjamin Daniels, Timothy Powell-Jackson, Christina Makungu, and Catherine Goodman. 2019. "How to do (or not to do)… using the standardized patient method to measure clinical quality of care in LMIC health facilities." *Health policy and planning* 34 (8): 625–634.

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

Kwan, Ada, Benjamin Daniels, Sofi Bergkvist, Veena Das, Madhukar Pai, and Jishnu Das. 2019. "Use of standardised patients for healthcare quality research in low-and middle-income countries." *BMJ global health* 4 (5): e001669.

Kwan, Ada, Benjamin Daniels, Vaibhav Saria, Srinath Satyanarayana, Ramnath Subbaraman, Andrew McDowell, Sofi Bergkvist, Ranendra K Das, Veena Das, Jishnu Das, et al. 2018. "Variations in the quality of tuberculosis care in urban India: A cross-sectional, standardized patient study in two cities." *PLoS medicine* 15 (9): e1002653.

Li, Jing, William H Dow, and Shachar Kariv. 2017. "Social preferences of future physicians." *Proceedings of the National Academy of Sciences* 114 (48): E10291–E10300.

Mills, Anne. 2014. "Health care systems in low-and middle-income countries." *New England Journal of Medicine* 370 (6): 552–557.

Mohanan, Manoj, Marcos Vera-Hernández, Veena Das, Soledad Giardili, Jeremy D Goldhaber-Fiebert, Tracy L Rabin, Sunil S Raj, Jeremy I Schwartz, and Aparna Seth. 2015. "The know-do gap in quality of health care for childhood diarrhea and pneumonia in rural India." *JAMA pediatrics* 169 (4): 349–357.

NMCP, KNBS, and ICF. *Kenya Malaria Indicator Survey 2015.*

Peabody, John W, Jeff Luck, Peter Glassman, Timothy R Dresselhaus, and Martin Lee. 2000. "Comparison of vignettes, standardized patients, and chart abstraction: a prospective validation study of 3 methods for measuring quality." *Jama* 283 (13): 1715–1722.

PharmAccess. 2017a. *SafeCare Basic Healthcare Standards Advanced Assessment Report.*

———. 2017b. *SafeCare Basic Healthcare Standards Basic Assessment Report.*

———. 2017c. *SafeCare Basic Healthcare Standards Quality Improvement Plan.*

———. 2017d. *SafeCare Basic Healthcare Standards Quality Improvement Plan.*

PharmAccess Group. 2017. *SafeCare Overview.*

Population Services Kenya. 2016. *AHME Business Assessment Example.*

Rasch, Georg. 1993. *Probabilistic models for some intelligence and attainment tests.* ERIC.

Rhodes, Karin V, and Franklin G Miller. 2012. "Simulated patient studies: an ethical analysis." *The milbank quarterly* 90 (4): 706–724.

Satyanarayana, S, R Subbaraman, P Shete, G Gore, J Das, A Cattamanchi, K Mayer, D Menzies, AD Harries, P Hopewell, et al. 2015. "Quality of tuberculosis care in India: a systematic review." *The international journal of tuberculosis and lung disease* 19 (7): 751–763.

Satyanarayana, Srinath, Ada Kwan, Benjamin Daniels, Ramnath Subbaraman, Andrew McDowell, Sofi Bergkvist, Ranendra K Das, Veena Das, Jishnu Das, and Madhukar Pai. 2016. "Use of standardised patients to assess antibiotic dispensing for tuberculosis by pharmacies in urban India: a cross-sectional study." *The Lancet Infectious Diseases* 16 (11): 1261–1268.

StataCorp, LLC. 2020. "Stata statistical software: release 16. College Station, TX."

Stiglitz, Joseph E. 1973. "Approaches to the Economics of Discrimination." *The American Economic Review* 63 (2): 287–295.

Suchman, Lauren, Dominic Montagu, and C. Avery Seefeld. 2020. "African Health Markets for Equity Qualitative Evaluation Final Report, 2013 – 2019." Accessed: July 17, 2020. https://globalhealthsciences.ucsf.edu/sites/globalhealthsciences.ucsf.edu/files/pub/african-health-markets-for-equity.pdf.

Sylvia, Sean, Yaojiang Shi, Hao Xue, Xin Tian, Huan Wang, Qingmei Liu, Alexis Medina, and Scott Rozelle. 2014. "Survey using incognito standardized patients shows poor quality care in China's rural clinics." *Health policy and planning* 30 (3): 322–333.

Sylvia, Sean, Hao Xue, Chengchao Zhou, Yaojiang Shi, Hongmei Yi, Huan Zhou, Scott Rozelle, Madhukar Pai, and Jishnu Das. 2017. "Tuberculosis detection and the challenges of integrated care in rural China: A cross-sectional standardized patient study." *PLoS medicine* 14 (10): e1002405.

Wang, Fahui, and Wei Luo. 2005. "Assessing spatial and nonspatial factors for healthcare access: towards an integrated approach to defining health professional shortage areas." *Health & place* 11 (2): 131–146.

Wilson, Nicholas. 2020. "If You Build It Nearby and Keep the Price Low, Will They Come? Large-Scale Evidence on the Relative Roles of Access and Quality in Use of Healthcare in Low-and Middle-Income Countries." *South African Journal of Economics* 88 (1): 3–20.

World Bank. 2008. *Kenya Poverty and Inequality Assessment.*

World Bank Group. 2020 (accessed April 21, 2020). *Poverty and Equity Data Portal.* `http://povertydata.worldbank.org/poverty/category/LIC`.

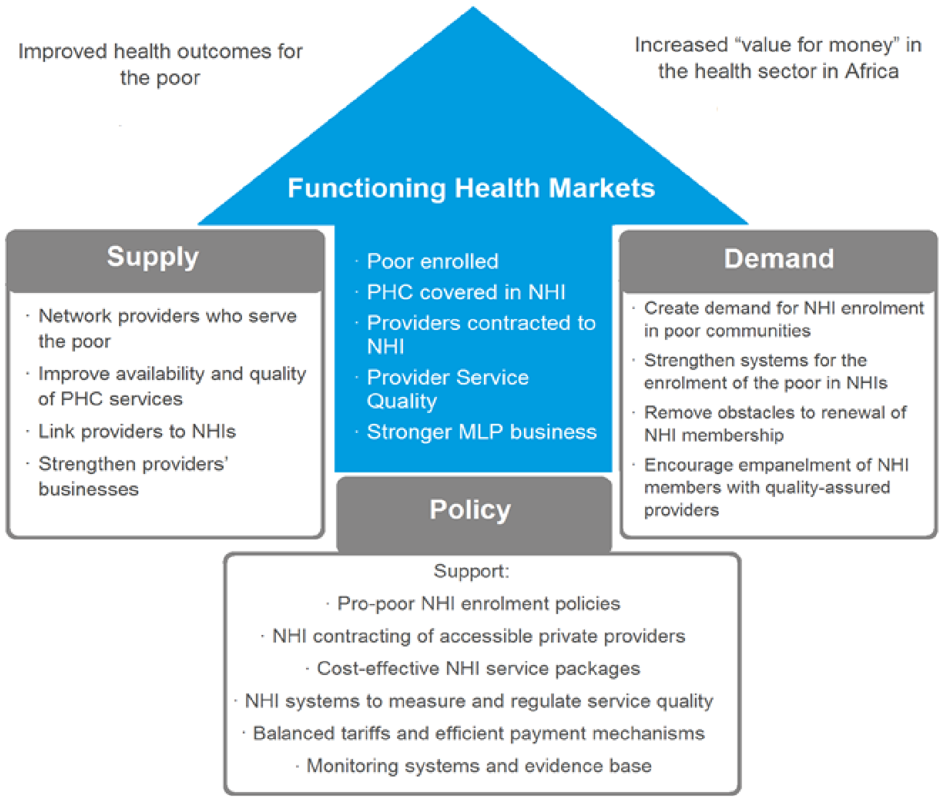
———. 2017 (accessed April 21, 2020). *Service Delivery Indicators 2017.* `www.sdindicators.org`.

World Health Organization. 2017. "Tracking universal health coverage: 2017 global monitoring report."

Figures and Tables

Figure 2.1. AHME Theory of Change

The design of the AHME program rested on the belief that supply, demand, and policy needed to be addressed together to create a health market that functions well for the poor.



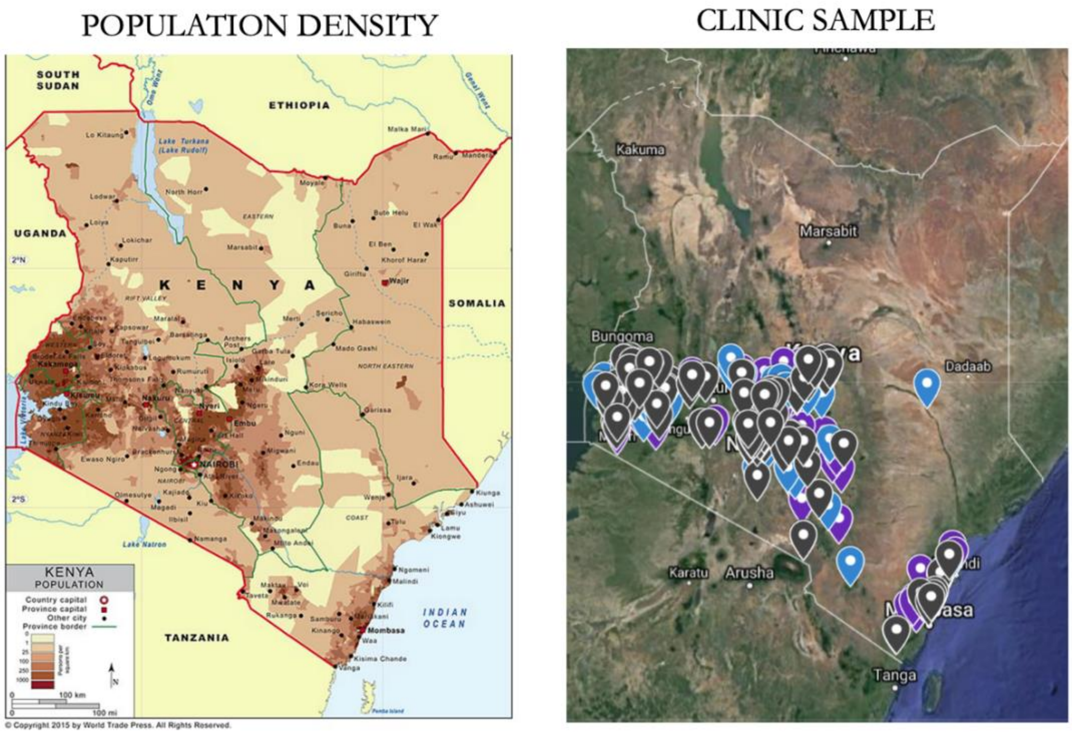
Note: MLP refers to mid-level provider; NHI refers to national health insurance; PHC refers to primary health care.

Figure 3.1 Quality of Care Conceptual Framework Based on the Donabedian Quality of Care Framework

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

Note: The depicted framework is based on Donabedian (1966, 1978, 1988).

Figure 4.1 Map: Kenya Population Density and AHME Evaluation Clinic Sample



Note: Population density of Kenya on left; AHME impact evaluation clinic sample on the right (control clinics in grey, and treatment clinics in blue and purple.

Figure 4.2 AHME Evaluation Endline Surveys

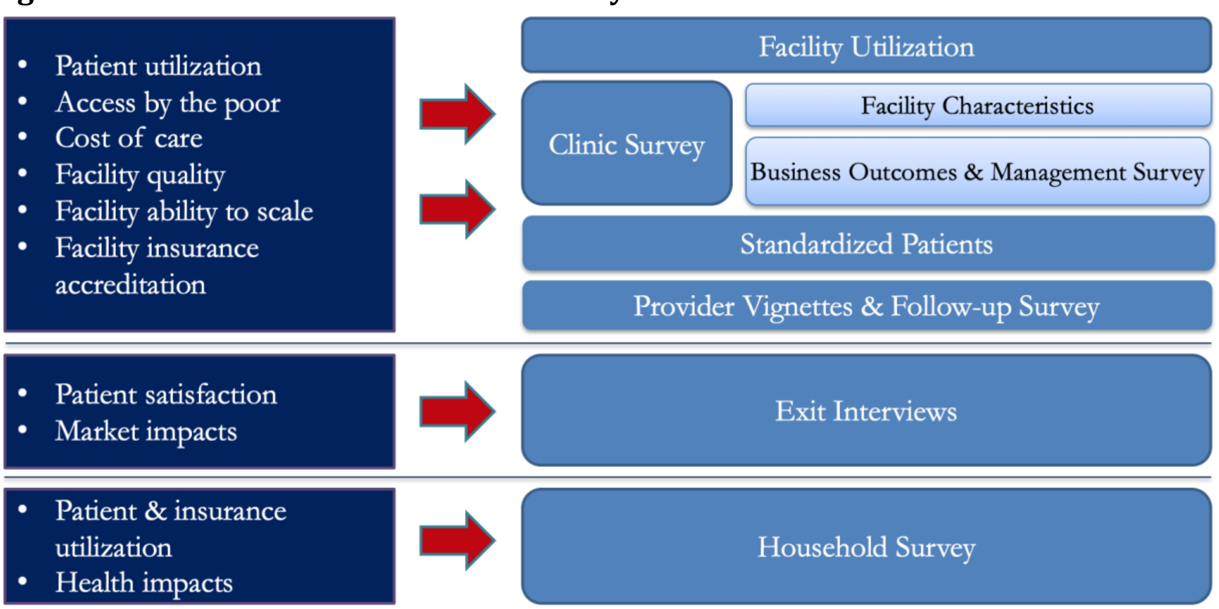


Table 4.1. Balance of Clinic Manager Characteristics at Endline by AHME Treatment

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

Note: Proportions reported, unless otherwise noted.

Table 4.2. Balance of Baseline Clinic Characteristics at Endline among Consented

	Control (n=91)		Treatment (n=108)		
	Mean	(SE)	Mean	(SE)	Difference (t-test)
Number of patients (last week)	95.6	(9.13)	87.0	(7.10)	8.52
Ownership (years)	9.86	(0.68)	10.6	(0.59)	-0.78
Provides ANC	0.67	(0.05)	0.67	(0.05)	0.00
Provides labor and delivery	0.41	(0.05)	0.40	(0.05)	0.01
Provides PNC	0.62	(0.05)	0.64	(0.05)	-0.02
Provides child immunization	0.46	(0.05)	0.30	(0.04)	0.16**
Provides well-baby check-ups	0.73	(0.05)	0.62	(0.05)	0.12*
Provides TB treatment (adults)	0.06	(0.02)	0.10	(0.03)	-0.04
Provides inpatient services	0.24	(0.04)	0.18	(0.04)	0.06
Share of family planning clients	0.17	(0.01)	0.16	(0.01)	0.01
Unit cost per client	13.9	(1.48)	13.6	(1.26)	0.28
Profit margin (average 6 months)	0.37	(0.03)	0.39	(0.02)	-0.02
Empaneled with NHIF	0.08	(0.03)	0.14	(0.03)	-0.06
Businesses pay for their employees to receive health care from this clinic	0.15	(0.04)	0.20	(0.04)	-0.05
Facility sells medicines to the public	0.48	(0.05)	0.51	(0.05)	-0.03
F-test of joint significance (F-statistic)					1.50
F-test, number of observations					199

Note: Proportions reported, unless otherwise noted. Standard errors in parentheses. The value displayed for t-tests 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. NHIF refers to National Hospital Insurance Fund; TB refers to tuberculosis; ANC refers to antenatal care; PNC refers to prenatal care.

Table 4.3. Comparison of AHME and DHS Household Populations

Household characteristic	AHME		DHS	
	Mean	(SD)	Mean	(SD)
Urban setting	0.42	(0.49)	0.41	(0.49)
Male head of household	0.89	(0.31)	0.64	(0.47)
Any education (head of household)	0.96	(0.20)	0.87	(0.34)
Finished floor	0.62	(0.48)	0.58	(0.49)
Improved water source	0.78	(0.41)	0.70	(0.46)
Improved sanitation (unshared)	0.34	(0.47)	0.28	(0.45)
	Median	IQR	Median	IQR
Age (years) of head of household	35	(30-42)	39	(30-53)
Household family members	5	(4-6)	3	(2-5)

Note: AHME refers to African Health Markets for Equity; DHS refers to Demographic and Health Surveys; IQR refers to Interquartile Range; SD refers to standard deviation.

Table 4.4. Balance of Baseline Household Characteristics among Analytic Sample

	(1)		(2)		(3)	(4)
	AHME Treatment		AHME Control		Diff.	F- stat
Variable	Mean	(SE)	Mean	(SE)		
Demographics						
Urban setting	0.44	(0.05)	0.40	(0.06)	0.04	
Number of HH members	5.07	(0.11)	5.10	(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)	0.90	(0.01)	-0.01	1.05
Age of head of HH	37.69	(0.58)	37.06	(0.64)	0.63	
Head of HH's years of schooling	11.02	(0.27)	10.28	(0.32)	0.74*	
Expectant mother in HH	0.09	(0.01)	0.12	(0.01)	-0.02	
Infrastructure						
Improved sanitation	0.37	(0.03)	0.30	(0.03)	0.07*	
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*	1.23
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						
Wealth score (DHS)	0.30	(0.07)	0.18	(0.09)	0.13	1.28
Clinic Experience						
Attention received[a]	0.83	(0.02)	0.87	(0.02)	-0.03	
Care received[a]	0.81	(0.02)	0.84	(0.02)	-0.03	
Wait time[a]	0.38	(0.03)	0.40	(0.03)	-0.02	0.91
Clinic cost[a]	0.28	(0.05)	0.35	(0.08)	-0.07	
Transportation cost in KSH	52.73	(4.28)	62.81	(7.00)	-10.08	
Overall experience[a]	0.83	(0.02)	0.86	(0.02)	-0.02	
Insurance						
Any HH member has insurance	0.42	(0.03)	0.34	(0.03)	0.08*	
Total HH members with insurance	1.44	(0.12)	1.25	(0.10)	0.19	1.80
Share of HH members with insurance	0.32	(0.03)	0.26	(0.02)	0.06*	

(Table continued on next page)

(Table continued from previous page)

Illness and Utilization [b]						
Child in household had fever	0.22	(0.02)	0.24	(0.02)	-0.03	1.50
Sought fever care	0.32	(0.03)	0.36	(0.03)	-0.04	
Child in HH: upper respiratory illness	0.24	(0.02)	0.20	(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	

*** p<0.01, ** p<0.05, * p<0.1. Robust standard errors (SE) in parentheses, clustered by index facility. Column (1)–(2) is the magnitude of treatment mean minus control mean, with asterisk from p-values from a two-sided t-test. F-statistics for groups of baseline variables are displayed in column 4. All missing values in balance variables are treated as zero. HH refers to household; KSH refers to Kenya shillings.

a. Ratings on a 1–5 Likert scale, with 1=worst/lowest, 5=best/highest, were recoded to indicator variables where 1=4,5 and 0=1,2,3.

b. 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 (SD) of 1 within each group.

Table 4.5. Balance of Analytic Clinic Sample for Standardized Patient (SP) Data

	(1)	(2)	(3)
	SP Sample (n=211)		
	Control (n=97)	Treatment (n=114)	
	Mean/SE	Mean/SE	Difference
Number of patients (last week)	97.29	85.30	11.98
	[8.53]	[6.76]	
Ownership (years)	10.10	10.08	0.02
	[0.70]	[0.59]	
Provides ANC	0.70	0.69	0.01
	[0.05]	[0.04]	
Provides labour and delivery	0.42	0.40	0.02
	[0.05]	[0.05]	
Provides PNC	0.62	0.65	-0.04
	[0.05]	[0.04]	
Provides child immunization	0.48	0.33	0.15**
	[0.05]	[0.04]	
Provides well-baby check-ups	0.74	0.61	0.13**
	[0.04]	[0.05]	
Provides TB treatment (adults)	0.08	0.10	-0.02
	[0.03]	[0.03]	
Provides inpatient services	0.24	0.19	0.05
	[0.04]	[0.04]	
Share of family planning clients	0.17	0.16	0.01
	[0.01]	[0.01]	
Average cost per client	796.15	29.55	766.60
	[735.86]	[6.78]	
Average revenue per client	1006.85	45.23	961.62
	[919.72]	[9.17]	
Profit margin (average 6 months)	0.38	0.41	-0.03
	[0.03]	[0.02]	
Profit margin (last month)	0.34	0.39	-0.04
	[0.03]	[0.02]	
Has NHIF	0.11	0.15	-0.04
	[0.03]	[0.03]	
Business paid for employees	0.18	0.21	-0.04
	[0.04]	[0.04]	
Facility sells medicines to the public	0.49	0.50	-0.01
	[0.05]	[0.05]	
F-test of joint orthogonality (F-stat)			1.37
F-test, number of observations			211

Note: Standard errors in brackets. The value displayed for t-tests 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. NHIF refers to National Hospital Insurance Fund; TB refers to tuberculosis; ANC refers to antenatal care; PNC refers to prenatal care.

Table 5.1. Structural Quality Measurement: Differences among Implementing Partners' (PSK, MSK, SafeCare) Internal Monitoring Indicators

	PSK	MSK	SafeCare
Number of clinics	37	56	109
Number of indicators	1,591 across 9 categories	402 across 3 categories	1,185 across 13 categories
Variable measurement[a]	Binary	Ordinal	Ordinal
Assessment date range	Sep 2016–April 2018	Feb 2016–Jun 2018	May 2012–Sep 2018
No. of assessments per clinic	1–2	1–4	1–3

Note: MSK refers to Marie Scopes Kenya; PSK refers to Population Services Kenya.
a. MSK scored each indicator as not in place, partially achieved, or fully achieved. PSK scored each indicator as yes/pass or no/fail. SafeCare scored each indicator as not, partially, or fully compliant.

Figure 5.1. Structural Quality Measurement: Criteria Used to Reduce the Number of Indicators

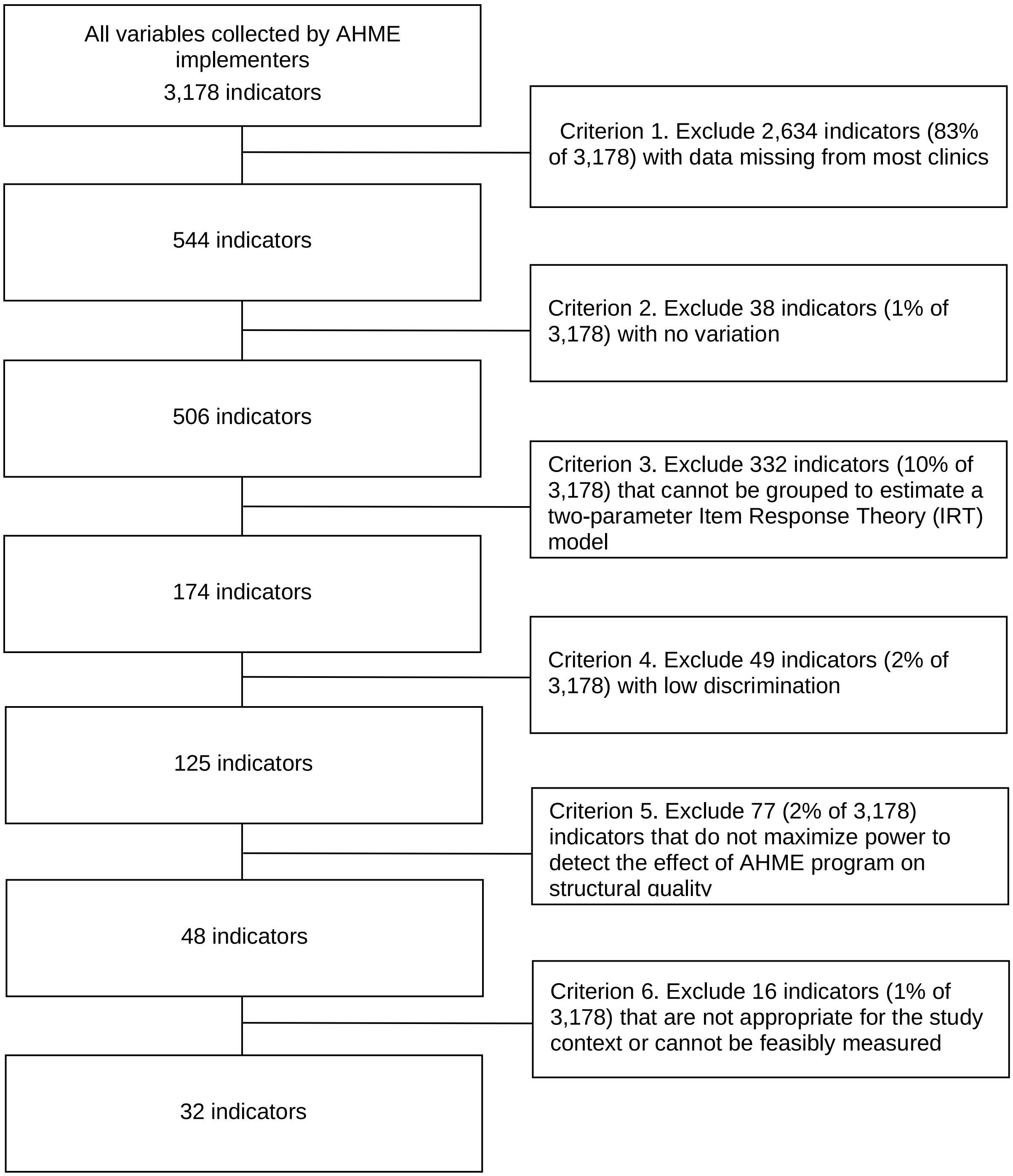


Table 5.2. Structural Quality Measurement: Implementers' Internal Monitoring Indicators Incorporated into AHME Endline Clinic Survey, by Dimension of Structural Quality

Implementer	Structural quality dimension (13)	Monitoring indicator (32, total)
PSK	IMCI clinical consultation and technical support	1. Provider asks if the child is coughing. Asks about frequency and duration of cough if the child is reported to be coughing.
		2. Provider asks if there is difficulty in breathing, fast breathing, wheezing.
		3. Provider asks whether the mother or other family members are coughing.
		4. Provider ask for HIV positivity in mother or child.
	Work environment	5. Professional license (doctors, nurses/midwives, CHEWs, etc.)
		6. Availability of IEC materials
MSK	General client-focused care	7. Consultation rooms are clean, spacious, and allow for privacy.
		8. Procedure rooms are clean.
		9. Procedure rooms have sufficient natural or electrical light, with back-up for outages.
		10. Drinks (water is sufficient) are available in recovery and in waiting areas.
	Clinical governance	11. Action plans developed from the last Clinical Quality Internal Audit are available and have been/are being acted upon, as evidenced by review of status of each action point.
		12. Training plans for clinical staff are informed by needs identified through continuous supportive supervisions.
		13. Dissemination of investigation outcomes, follow-up of action plans, and learning points from incidents is documented in staff meeting minutes.
	Supplies management and product quality	14. There is a designated staff who is responsible for stock management for the facility/team.
		15. Responsible stock officer has the "standard product sub-list" for their facility appropriate to their channel, and can state target minimum/maximum stock, normal frequency of orders, and date of next order.
		16. For the given product above, the actual stock level is within the target minimum and maximum quantities.

(Table continued on next page)

(Table continued from previous page)

SafeCare	Human resource management	17. Each employee has a written job description/performance agreement, which defines their responsibilities.
	Governance and management	18. Documents describe governance accountability and responsibilities. 19. The system includes safe handling, storing, and disposing of different types of waste. 20. Each patient has access to a nurse call system at all times.
	Medication management	21. The consultation rooms are clean, well ventilated, well maintained, and adequately equipped. 22. There is a list of the medications, stocked in the organization, or readily available from outside sources. 23. Patients who require early attention are identified (e.g. the very frail or ill, or women in an advanced stage of pregnancy).
	Non-clinical patient care	24. The patient's need for privacy is protected during all examinations, procedures and treatments. 25. Organizational policy regarding patient and family rights is implemented.
	Clinical management	26. The health facility uses a health information system that facilitates the collection and utilization of data. 27. An on-call roster is available for after-hour, weekend, and holiday emergency coverage (e.g. for infectious diseases).
	Clinical processes and infrastructure	28. There is an organized process for referring patients. 29. The available supplies, reagents, chemicals, and kits are sufficient for projected activities.
	Primary health care services	30. Written guidelines for providing primary emergency services are available and followed. 31. The health facility has a procedure, which is implemented, when others have to grant informed consent.
	IMCI clinical consultation and technical support	32. Guidelines for postnatal care are available and followed.

Note: CHEW refers to community health extension workers; HIV refers tohuman immunodeficiency virus; IEC refers to information, education, and communication; IMCI refers to Integrated Management of Childhood Illness; MSK refers to Marie Stopes Kenya; PSK refers to Population Services Kenya.

Table 5.3. Structural Quality Measurement: Compliance Measures

Dimension of structural quality	Indicator
Structures to facilitate interpersonal care (16 indicators, including material/ human resources and organizational structures that facilitate the exchange through which patients and providers share information and preferences)	At least two color-coded bins with matching bags
	Dedicated container and separate disposal of sharps and bio/chem/pharm waste
	Disinfected smell where health services are offered
	Drinking water in the waiting area and recovery room
	Exits, corridors, and rooms free of hazards
	Facility keeps record of incidents and takes action
	Facility provides IEC materials
	Floors and walls are in good condition
	Floors and walls have good hygiene
	Governing board exists and responsibilities are documented, including community accountability
	Informed consent for some health services
	No cobwebs, etc. on the ceiling/walls
	Require a note when referring and receiving referrals
	Sufficient natural or electrical light
	Visible patient rights chart
	Visual and auditory privacy in consultation space
Structures to facilitate technical care (17 indicators, including material/ human resources and organizational structures that facilitate the provider's recommendation of appropriate strategies of care and implementation of those strategies)	Designated staff for stock management
	Electricity source in case of emergency
	Facility has training plans that are based at least partially on direct observation of clinical duties
	Facility keeps staff files with credentials
	Gloves are available
	Information technology or smart phone used in administration/operation of clinic
	License to have private medical practice
	List of referral sites to be consulted
	Master list of health inputs specifies minimum/maximum stock and order frequency
	Master list of health inputs to be kept in stock
	No shortage of essential medicines/products in the last year
	Observation room with beds
	On-call roster of health staff
	Performs internal clinical quality audit, creates action plan, and acts upon it
	Printed guidelines for postnatal care
	Printed guidelines for primary emergency services
	Written job description or performance agreement for employees

Note: IEC refers to information, education, and communication.

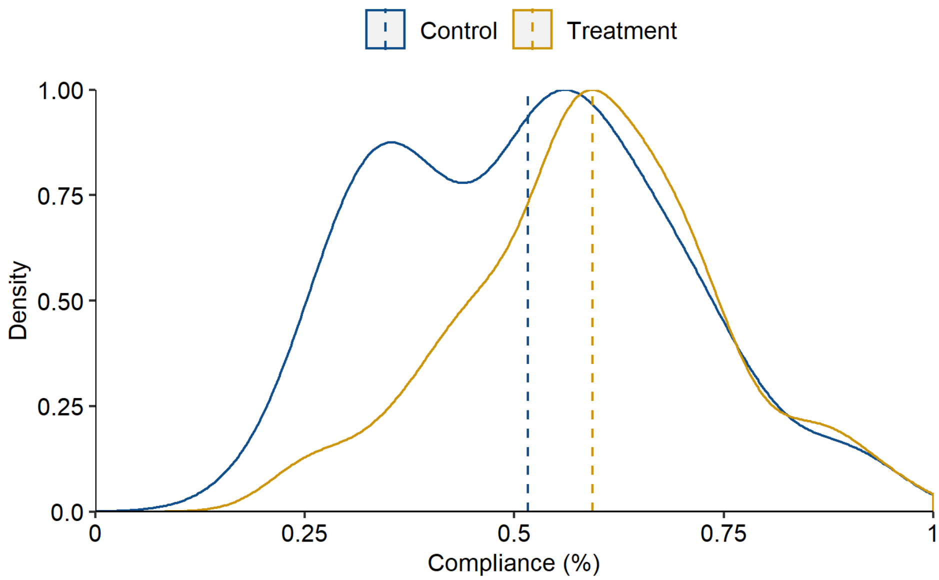
Table 5.4. AHME Effects on Structural Quality Compliance by Dimension

	Structural quality		
	To facilitate interpersonal care	To facilitate technical care	Overall
AHME treatment			
Coefficient	0.079	0.075	0.077
Standard error	(0.024)	(0.027)	(0.023)
p-value	[0.001]	[0.003]	[<0.001]
Mean control group	0.508	0.525	0.516
Number of observations	199	199	199

Note: All models are linear regressions of the measure of structural quality compliance on a binary variable indicating whether the clinic was assigned the AHME intervention. Compliance takes a value of 0 to 1 and represents the share of associated indicators met in the clinic. The coefficient on AHME treatment is the percentage point difference in compliance relative to control clinics. One-sided p-values are in brackets (H_0: $\beta \leq 0$).

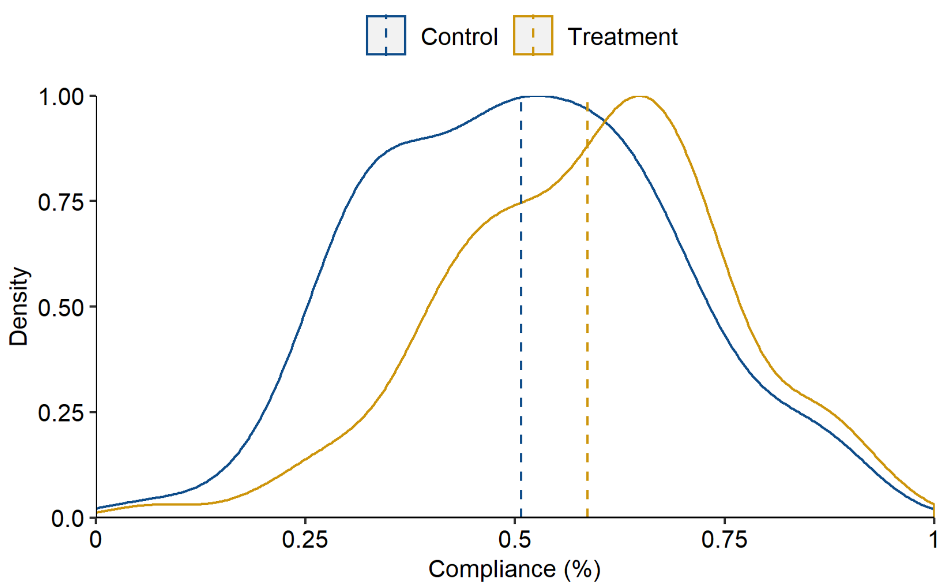
Figure 5.2. Conditional Density of Structural Quality Compliance by Dimension
The distribution of structural quality compliance narrows slightly among treated clinics compared to control clinics, particularly for overall structural quality compliance.

a. Overall structural quality



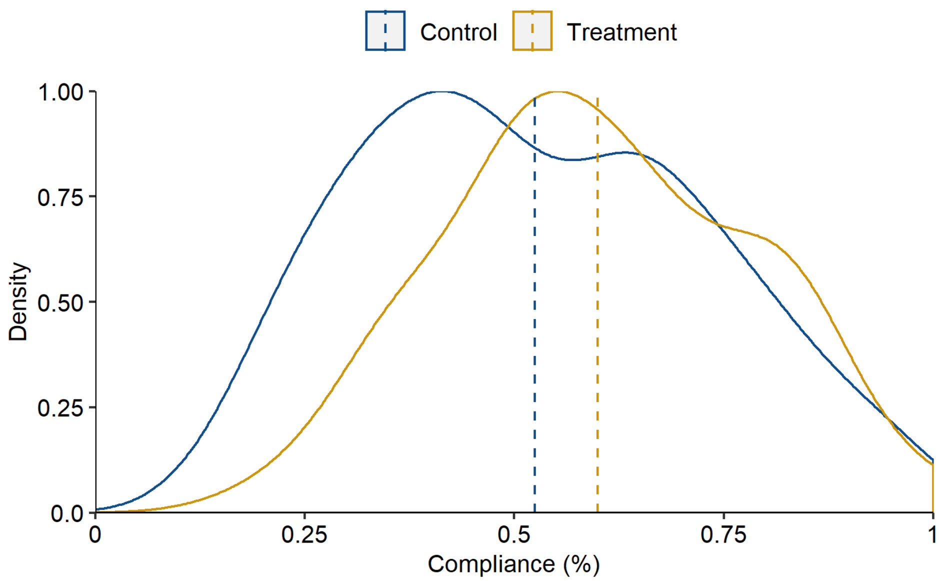
Note: Solid lines are the kernel density of compliance. Dashed lines are conditional means of compliance. Compliance takes a value of 0 to 1 and represents the portion of associated indicators met in the clinic. Density and means are calculated from a sample of 108 treatment clinics and 91 control clinics.

b. Structures to facilitate interpersonal care



Note: Solid lines are the kernel density of compliance. Dashed lines are conditional means of compliance. Compliance takes a value of 0 to 1 and represents the portion of associated indicators met in the clinic. Density and means are calculated from a sample of 108 treatment clinics and 91 control clinics.

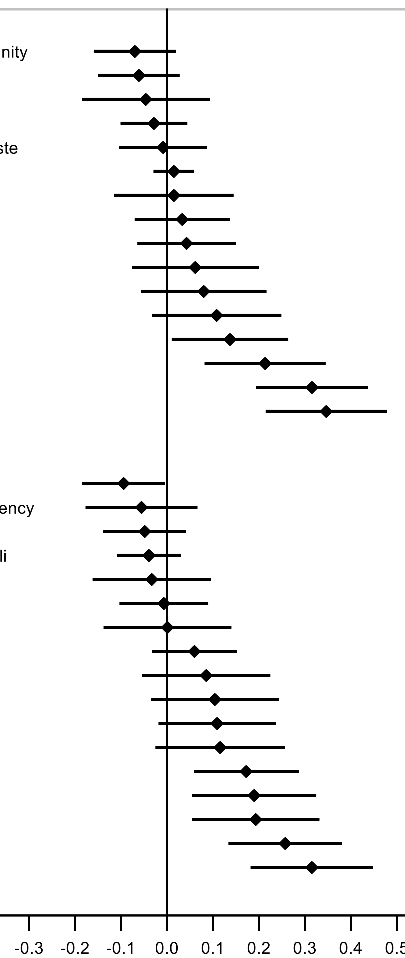
c. Structures to facilitate technical care



Note: Solid lines are the kernel density of compliance. Dashed lines are conditional means of compliance. Compliance takes a value of 0 to 1 and represents the portion of associated indicators met in the clinic. Density and means are calculated from a sample of 108 treatment clinics and 91 control clinics.

Figure 5.3. AHME Effects on Structural Quality Compliance, by Indicator
Compliance varied widely across the 33 individual structural quality indicators.

Structural quality indicators	N	Control mean
Interpersonal		
Governing board exists and responsibilities are documented, including community	191	0.148
Exits, corridors, and rooms free of hazards	198	0.922
Floors and walls are in good condition	198	0.611
Visual and auditory privacy in consultation space	197	0.944
Dedicated container and separate disposal of sharps and bio/chem/pharm waste	179	0.122
Sufficient natural or electrical light	198	0.967
No cobwebs etc. on the ceiling/walls	198	0.689
Disinfected smell where health services are offered	199	0.143
Drinking water in the waiting area and recovery room	127	0.081
At least two color-coded bins with matching bags	199	0.374
Floors and walls have good hygiene	198	0.578
Facility keeps record of incidents and takes action	196	0.411
Require a note when referring and receiving referrals	181	0.675
Facility provides IEC materials	199	0.527
Informed consent for some health services	199	0.527
Visible patient rights chart	198	0.256
Technical		
License to have private medical practice	197	0.934
Master list of health inputs specifies minimum/maximum stock and order frequency	198	0.278
Gloves are available	186	0.919
Information technology or smart phone used in administration/operation of facili	199	0.956
On-call roster of health staff	198	0.311
Designated staff for stock management	199	0.868
Master list of health inputs to be kept in stock	199	0.582
Facility has training plans that are based at least partially on direct observat	196	0.090
Electricity source in case of emergency	199	0.516
No shortage of essential medicines/products in the last year	197	0.522
Printed guidelines for postnatal care	197	0.231
Performs internal clinical quality audit, creates action plan, and acts upon it	187	0.321
Observation room with beds	198	0.689
Printed guidelines for primary emergency services	197	0.278
Written job description or performance agreement for employees	197	0.433
Facility keeps staff files with credentials	194	0.575
List of referral sites to be consulted	198	0.389



Note: All models are linear regressions of a structural quality indicator on a binary variable indicating whether the clinic was assigned the AHME intervention. Structural quality indicators take a value of 0 or 1 and represents whether the indicator was met in the clinic. The coefficient on AHME treatment is the difference in the probability that a treated clinic complies with an indicator relative to control clinics. Black diamonds represent the coefficients on AHME treatment from separate models. Black bars represent 90% confidence intervals. IEC refers to information, education, and communication.

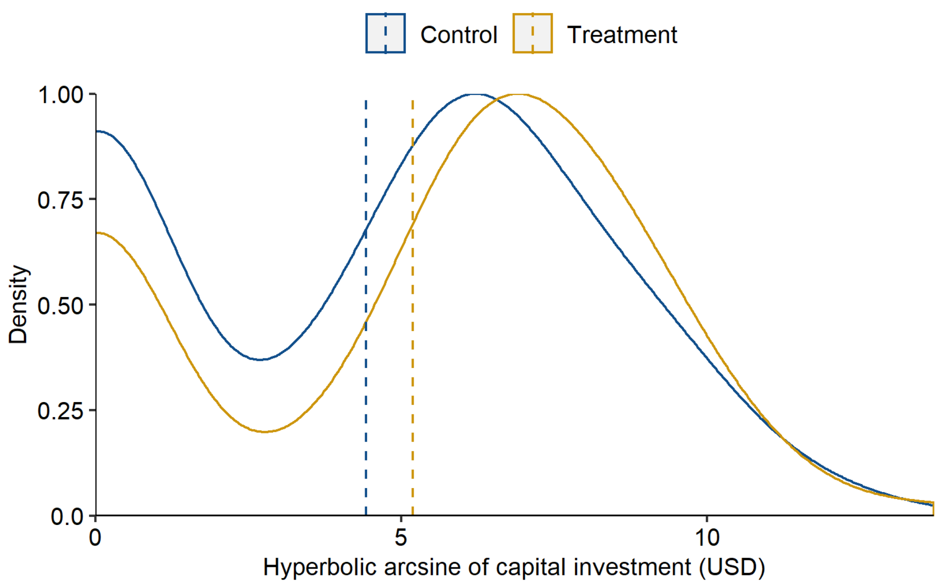
Table 5.5. AHME Effects on Capital Investments in Amenities and Equipment

	Investment					
	Total	Amenities	Diagnostic equipment	Laboratory equipment	IT equipment	Medical equipment
AHME treatment						
Coefficient	0.765	0.172	-0.195	-0.225	0.653	0.866
Standard error	(0.534)	(0.287)	(0.333)	(0.448)	(0.490)	(0.494)
p-value	[0.076]	[0.275]	[0.720]	[0.692]	[0.092]	[0.041]
Mean of control	2,539	1,161	76	689	274	308
Observations	186	187	187	187	187	187

Note: All models are linear regressions of investments on a binary variable indicating whether the clinic was assigned the AHME intervention. Investments were measured in their resale value (in Kenyan shillings, KSH), converted to United States dollars (USD), winsorized at the 1st and 99th percentiles, and transformed to hyperbolic arcsine. Amenities include air conditioning, room heaters, a water dispenser, and lockers. Diagnostic equipment includes microscope, glucometer, sphygmomanometer, and ophthalmoscope. Laboratory equipment includes biochemistry analyzer, ELISA reader, centrifuge, hemogram machine, ultrasound machine, and x-ray machine. Information technology (IT) equipment includes computer, printer, and projector. Medical equipment includes beds, exploration tables, negatoscope, oxygen machine, refrigerator, sterilizer, and ventilator machine. The coefficient is interpreted as the percent change relative to the mean of the control group (in US dollars). One-sided p-values are in brackets (H_0: $\beta \leq 0$).

Figure 5.4. Conditional Density of Capital Investments

There is a wide and bimodal distribution among control and treatment clinics of structural quality compliance with respect to total capital investments.



Note: Solid lines are the kernel density of compliance. Dashed lines are conditional means of capital investments. Investments were measured in their resale value (in Kenyan shillings, KSH), converted to United States dollars (USD), winsorized at the 1st and 99th percentiles, and transformed to hyperbolic arcsine. Density and means are calculated from a sample of 103 treatment clinics and 83 control clinics.

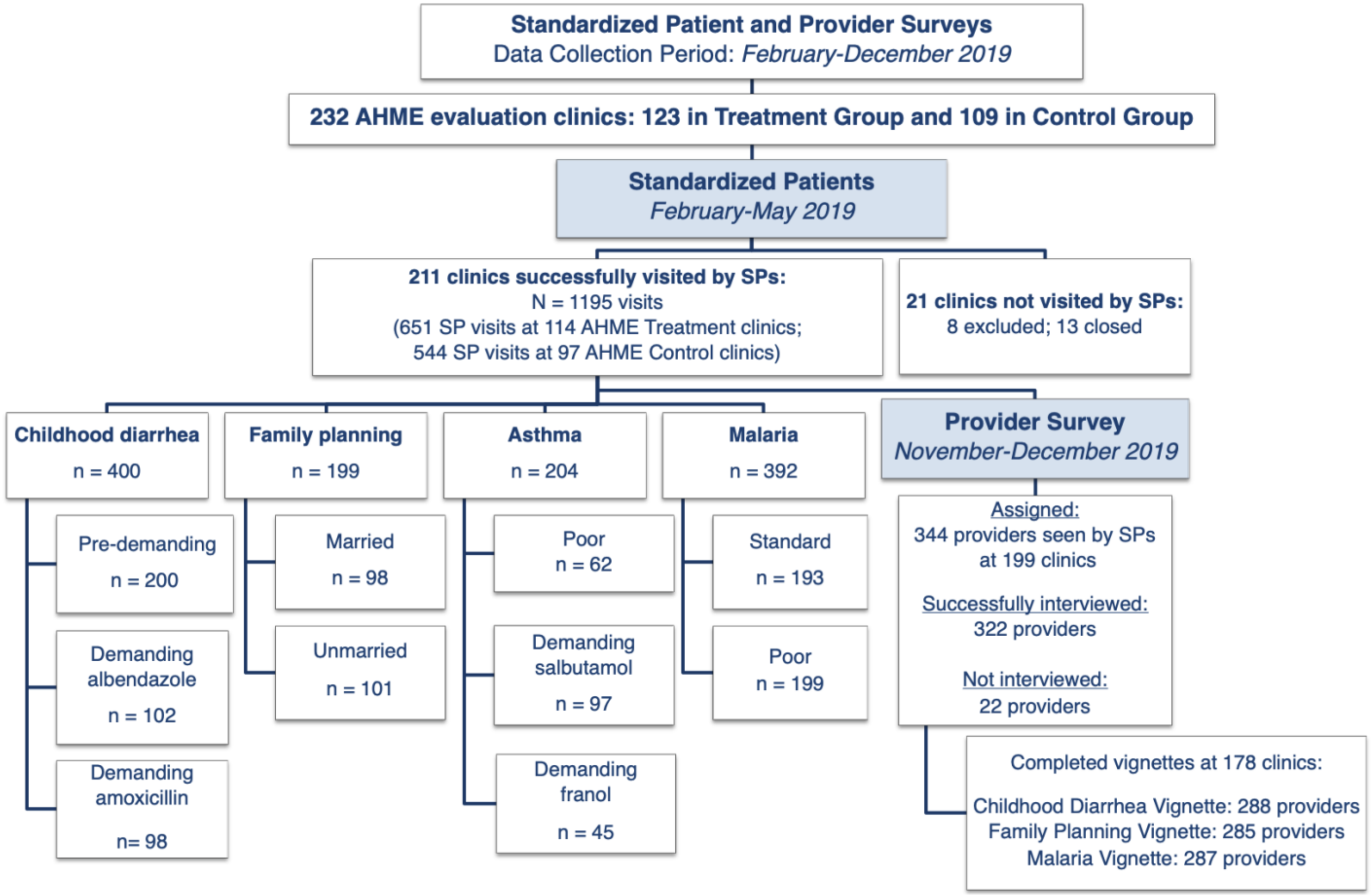
Table 6.1. Process Quality Measurement: Description of Standardized Patient (SP) and Vignette Case Scenarios

Case	SP	Vig-nette	Case description	SP opening statement	SP experiments
Childhood diarrhea	✓	✓	A 28-year-old mother comes to the clinic; her 1.5-year-old child is at home sick with diarrhea.	"My child has been having diarrhea."	Pre-demanding vs. Demanding albendazole vs. Demanding amoxicillin
Family planning	✓	✓	A 27-year-old married female seeks family planning advice since she does not want to have children for another 2 years.	"I do not wish to have any other children for 2 years, and I would like to know what methods are possible."	Married vs. 25-year-old unmarried female
Asthma	✓	X	A 24/25-year-old female/male presents with great difficulty breathing the previous night.	"Doctor, last night I had a lot of difficulty with breathing."	Standard case vs. Cannot afford more than KSH 300 ("Poor")
Malaria	✓	✓	A 28-year-old female or male who has recently traveled to a malaria-endemic area presents with fever and headache and thinks s/he has malaria.	"Doctor, I think I have malaria."	Standard case vs. Cannot afford more than KSH 300 ("Poor")

Note: Checks under the standardized patient (SP) column refer to whether the case was implemented for SP endline surveys. Under the vignette column, checks refer to whether the case was implemented as a clinical vignette in the endline provider survey. KSH refers to Kenyan shillings (where 100 KSH = 1 United States Dollar).

Figure 6.1. Process Quality Measurement: AHME Standardized Patient Survey and Provider Survey Sampling

The standardized patient survey conducted as walk-in visits identified respondents for the provider survey administered 6 months later.



Note: Three vignettes cases were conducted to assess provider knowledge for childhood diarrhea, family planning, and malaria cases with similar presentation as the standardized patient (SP) cases. For SP cases: all clinics were assigned 1 childhood diarrhea case, 1 family planning case, 1 asthma case, and 2 malaria cases. SP experiments for childhood diarrhea, family planning, and asthma (demanding albendazole or amoxicillin; married or unmarried; poor or demanding salbutamol or franol) were randomly assigned (fully independent of AHME treatment assignment) to clinics. All clinics were assigned to receive both the standard and poor SP experiment for the malaria case. For the childhood diarrhea SP case, demanding experiment occurs at the end of interaction (n=200), and any actions conducted as pre-demanding are analyzed as the standard case for the SP experiment (n=200); therefore, we analyze n=400 childhood diarrhea SP visits. During the provider survey, the AHME clinic sample was restricted to n=199 clinics based on response rates from the AHME clinic survey (conducted between February-May 2019).

Table 6.2. Process Quality: Main Outcomes by SP and Vignette Case Scenarios

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

Note: Correct case management is a binary variable: 1=yes if the provider performed the actions listed; 0 if otherwise. Each valid test, unnecessary test, efficacious medicine, and non-efficacious medicine is a binary variable 1=yes if the provider ordered/prescribed; 0=no if otherwise. FP refers to family planning; ORS refers to oral rehydration salts; SP refers to standardized patient.

Table 6.3. Process Quality: Correct Case Management Components for Family Planning Case

Case	Components of correct case management	Elements of component
Family planning	Asked any FP method history	History: Are you currently using an FP method? History: Have you previously used an FP method?
	Asked any obstetric history	History: Have you been previously pregnant? or How many times have you been pregnant? History: When was your last pregnancy? History: What happened to your previous pregnancy? or Have you had a miscarriage or stillborn child? History: How many children do you have?
	Ruled out pregnancy	History: When was your last monthly period? History: Have you had any unprotected sex? Lab test: Pregnancy test
	Asked preferred FP method	History: When would you like to get pregnant? Other: Which FP method would you prefer?

Note: FP refers to family planning. Each correct case management component for family planning is a binary variable that is 1=yes if any one of the elements of the component is asked or performed; 0=no if otherwise.

Table 6.4. Process Quality: Summary Statistics for SP Data

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
				Cases			
	Pooled N=1195	AHME control N=544	AHME treatment N=651	Diarrhea N=200	Family planning N=199	Malaria N=392	Asthma N=204
Providers by age group							
<30	0.176	0.166	0.184	0.175	0.173	0.182	0.167
30-50	0.576	0.608	0.548	0.603	0.569	0.567	0.530
>50	0.249	0.226	0.268	0.222	0.259	0.251	0.303
Provider is female	0.386	0.336	0.423	0.370	0.483	0.378	0.331
Avg. no. of clients in waiting room	1.527	1.708	1.376	1.550	1.623	1.523	1.397
Avg. minutes spent with provider	10.539	10.334	10.713	6.989	12.528	12.712	11.492
Avg. no. of history questions asked	5.855	5.833	5.874	4.455	6.658	7.730	4.471
Asked to return	0.403	0.383	0.419	0.295	0.619	0.509	0.250
Referred elsewhere	0.026	0.034	0.019	0.032	0.041	0.011	0.030
Correct diagnosis or suspicion	0.471	0.450	0.488	0.150		0.844	0.392
Correct case management	0.599	0.636	0.568	0.695	0.251	0.793	0.466
Any lab tests	0.332	0.364	0.306	0.130	0.055	0.827	0.054
Avg. no. of lab tests	0.500	0.540	0.465	0.255	0.065	1.207	0.054
Any unnecessary lab tests	0.141	0.154	0.131	0.100	0.020	0.298	0.054
Avg. no. of unnecessary lab tests	0.173	0.187	0.160	0.148	0.030	0.338	0.054
Avg. no. of meds	1.810	1.867	1.764	2.500	0.469	1.904	1.613
Avg. no. of non-efficacious meds	1.186	1.225	1.154	1.711	0.042	1.359	0.961
Knowledge of correct management							
Diarrhea	0.902	0.902	0.902	0.900	0.914	0.908	0.783
Family planning	0.120	0.091	0.142	0.117	0.109	0.128	0.130
Malaria	0.978	0.987	0.971	0.978	0.964	0.983	1.000

Note: SP refers to standardized patients. The table displays summary statistics are for all cases pooled in column 1; all cases pooled restricted to AHME control clinics only in column 2; all cases pooled restricted to AHME treatment clinics only in column 3; and by standardized patient (SP) case only in columns 4–7. All summary statistics are proportions unless variable is noted as average (avg.). Sample size for diarrhea interactions reflects post-demanding data. All summary statistics except knowledge of correct management for diarrhea, family planning, and malaria come from SP surveys. Knowledge of correct management is defined in the same way as correct case management and come from three vignettes administered in the provider survey. Vignette data are matched to SP data for each SP visit by provider seen by SP or a replacement for the sampled provider. Single observations for each provider in the provider survey may occur multiple times if SPs saw that provider multiple times. Provider age group is the estimated age group as perceived by the SP.

Table 6.5. AHME Effects on Correct Case Management: SP Regression Models

	(1) Diarrhea	(2) Family planning	(3) Asthma	(4) Malaria	(5) Pooled
AHME treatment					
Coefficient	-0.069	-0.084	-0.069	-0.092	-0.077
Standard Error	(0.048)	(0.060)	(0.069)	(0.049)	(0.033)
p-value	[0.155]	[0.165]	[0.324]	[0.058]	[0.021]
Mean control group	0.686	0.264	0.500	0.850	0.636
Observations	400	199	204	392	1195

Note: The table shows multivariate regressions using standardized patient (SP) data. Standard errors are in parentheses for models 1–4. Robust standard errors (in parentheses) are clustered at the clinic level for pooled model 5. All models contain SP fixed effects and contain covariates for the 0-1 AHME treatment indicator and a 0-1 indicator for each SP experiment (demanding, unmarried, poor). Model 5 contains case fixed effects. Correct case management is a 0-1 binary measure constructed specific to each scenario. For diarrhea SP data, correct case management = 1 is defined as whether the provider gave or advised on oral rehydration salts (ORS) or referred or asked the client to return; 0 otherwise. For family planning, SP data were coded as correctly managed = 1 if the provider performed all four of the following actions: asked any family planning history questions, asked obstetric history questions, ruled out pregnancy, and asked the client her preferred family planning method; 0 otherwise. For asthma, SP data were coded as correctly managed = 1 if the provider treated the case with an inhaler or bronchodilator; 0 otherwise. For malaria, SP data were coded as correctly managed = 1 if the provider ordered a malaria rapid diagnostic test (RDT) or a malaria microscopy test; 0 otherwise.

Table 6.6. AHME Effects on IMCI: Childhood Diarrhea SP Regression Models

	(1) Provider asked if child has cough or difficulty or rapid breathing	(2) Provider asked about breathing difficulty or rapid breathing	(3) Provider asked if anyone in family is coughing	(4) Provider asked about child's HIV status
AHME treatment				
Coefficient	0.015	0.012	0.000	-0.005
Standard Error	(0.034)	(0.012)	0.000	(0.005)
p-value	[0.667]	[0.317]	[.]	[0.318]
Mean control group	0.044	0.000	0.000	0.005
Observations	378	378	378	378

Note: IMCI refers to the Integrated Management of Childhood Illnesses approach. The table shows multivariate regressions using standardized patient (SP) data. Standard errors are in parentheses for models 1–4. Robust standard errors (in parentheses) are clustered at clinic level. Two-sided p-values in brackets. All models contain SP fixed effects and contain covariates for the 0-1 AHME treatment indicator and the childhood diarrhea SP experiment (demanding). Observations contain all successful and completed visits and do not include data of childhood diarrhea SP visits that resulted in immediate referral or the SP being turned away for child in absentia. The dependent variables for models 1–4 are binary outcomes for whether the provider asked the SP the history question. HIV refers to human immunodeficiency virus.

Table 6.7. AHME Effects on Any Unnecessary Lab Tests: SP Regression Models

	(1) Diarrhea	(2) Family planning	(3) Asthma	(4) Malaria	(5) Pooled
AHME treatment					
Coefficient	-0.070	-0.009	-0.028	0.044	-0.019
Standard Error	(0.029)	(0.021)	(0.031)	(0.054)	(0.024)
p-value	[0.015]	[0.656]	[0.366]	[0.415]	[0.423]
Mean control group	0.138	0.022	0.065	0.289	0.154
Observations	400	199	204	392	1195

Note: The table shows multivariate regressions using standardized patient (SP) data. Standard errors are in parentheses for models 1–4. Robust standard errors (in parentheses) are clustered at the clinic level for pooled model 5. All models contain SP fixed effects and control for the 0-1 AHME treatment indicator and a 0-1 indicator for each SP experiment (demanding, unmarried, poor). Model 5 contains case fixed effects. Any unnecessary lab tests is a 0-1 binary measure constructed specific to each scenario. For childhood diarrhea, SP data were coded as any unnecessary lab tests = 1 if providers ordered any lab tests excluding stool test; 0 otherwise. For family planning, SP data were coded as any unnecessary lab tests = 1 if providers ordered any lab tests excluding pregnancy test; 0 otherwise. For asthma, SP data were coded as any unnecessary lab tests = 1 if providers ordered any lab test; 0 otherwise. For malaria, SP data were coded as any unnecessary lab tests = 1 if providers ordered any lab test excluding malaria rapid diagnostic test (RDT), malaria microscopy, blood count, and brucellosis test; 0 otherwise.

Table 6.8. AHME Effects on Any Unnecessary Medicines: SP Regression Models

	(1) Diarrhea	(2) Family planning	(3) Asthma	(4) Malaria	(5) Pooled
AHME treatment					
Coefficient	-0.106	-0.043	-0.058	0.005	-0.050
Standard Error	(0.048)	(0.039)	(0.072)	(0.048)	(0.033)
p-value	[0.027]	[0.269]	[0.420]	[0.914]	[0.139]
Mean control group	0.723	0.088	0.554	0.717	0.586
Observations	400	199	204	392	1195

Note: The table shows multivariate regressions using standardized patient (SP) data. Standard errors are in parentheses for models 1–4. Robust standard errors (in parentheses) are clustered at the clinic level for pooled model 5. All models contain SP fixed effects and control for the 0-1 AHME treatment indicator and a 0-1 indicator for each SP experiment (demanding, unmarried, poor). Model 5 contains case fixed effects. Any unnecessary medicines is a 0-1 binary measure constructed specific to each scenario. For childhood diarrhea, SP data were coded as any unnecessary medicines = 1 if providers dispensed/prescribed any medicines excluding oral dehydration salts (ORS) or zinc; 0 otherwise. For family planning, SP data were coded as any unnecessary medicines = 1 if providers dispensed/prescribed any medicines excluding contraceptive pills; 0 otherwise. For asthma, SP data were coded as any unnecessary medicines = 1 if providers dispensed/ prescribed any medicines excluding inhaler or bronchodilators; 0 otherwise. For malaria, SP data were coded as any unnecessary medicines = 1 if providers dispensed/prescribed any medicines excluding artemether lumefantrine and paracetamol; 0 otherwise.

Table 6.9. AHME Effects on Family Planning (FP) Treatment: SP Regression Models

a. Correct management components

	(1) Asked family planning history	(2) Asked obstetric history	(3) Ruled out pregnancy	(4) Asked which FP method preferred
AHME treatment				
Coefficient	-0.001	-0.013	-0.163	-0.014
Standard Error	(0.052)	(0.072)	(0.072)	(0.063)
p-value	[0.984]	[0.860]	[0.024]	[0.829]
Unmarried				
Coefficient	-0.104	-0.043	-0.176	0.061
Standard Error	(0.062)	(0.086)	(0.085)	(0.075)
p-value	[0.096]	[0.619]	[0.041]	[0.416]
Mean control group	0.857	0.604	0.505	0.703
Observations	199	199	199	199

Note: The table shows multivariate regressions using standardized patient (SP) data. Each observation is one SP-provider visit at a different clinic. Standard errors are in parentheses. Two-sided p-values in brackets. All models contain SP fixed effects and control for the 0-1 AHME treatment indicator and the family planning SP experiment (unmarried). The dependent variable for model 1 is a 0-1 indicator for whether the provider asked the SP any history questions related to family planning. The dependent variable for model 2 is a 0-1 indicator for whether the provider asked the SP any questions related to obstetric history. The dependent variable for model 3 is a 0-1 indicator for whether the provider ruled out pregnancy: in other words, by asking when was your last monthly period or asking have you had any unprotected sex or ordering a pregnancy test. The dependent variable for model 4 is a 0-1 indicator for whether the provider asked the SP what family planning method she preferred.

b. Family planning counseling

	(1) Modern method: Side effects mentioned	(2) Modern method: Offered choice	(3) Traditional method: Side effects mentioned	(4) Traditional method: Suggested
AHME treatment				
Coefficient	0.068	0.033	0.075	-0.098
Standard Error	(0.059)	(0.064)	(0.040)	(0.044)
p-value	[0.251]	[0.604]	[0.060]	[0.027]
Unmarried				
Coefficient	0.000	0.126	0.060	0.030
Standard Error	(0.070)	(0.076)	(0.047)	(0.053)
p-value	[0.998]	[0.098]	[0.204]	[0.574]
Mean control group	0.758	0.648	0.055	0.165
Observations	199	199	199	199

Note: The table shows multivariate regressions using standardized patient (SP) data. Each observation is one SP-provider visit at a different index clinic. Standard errors are in parentheses. Two-sided p-values in brackets. All models contain SP fixed effects and control for the 0-1 AHME treatment indicator and the family planning SP experiment (unmarried). The dependent variable for model 1 is whether the provider mentioned side effects for modern family planning methods to the SP. The dependent variable for model 2 is whether the provider offered the SP a choice of modern family planning methods. The dependent variable for model 3 is whether the provider mentioned side effects for traditional family planning methods to the SP. The dependent variable for model 4 is whether the provider suggested a traditional family planning method to the SP.

Table 6.10. AHME Effects on Malaria Treatment: SP Regression Models

a. **Malaria diagnostics**

	(1)	(2)	(3)	(4)	(5)
	Temperature attempted with thermometer	Temperature taken by touch	Malaria test	Malaria rapid diagnostic test (RDT)	Malaria microscopy
AHME treatment					
Coefficient	-0.043	0.051	-0.084	-0.101	0.017
Standard Error	(0.066)	(0.030)	(0.048)	(0.062)	(0.065)
p-value	[0.512]	[0.091]	[0.085]	[0.102]	[0.790]
Poor					
Coefficient	-0.085	-0.007	-0.149	-0.023	-0.152
Standard Error	(0.063)	(0.032)	(0.054)	(0.064)	(0.069)
p-value	[0.182]	[0.816]	[0.007]	[0.724]	[0.029]
Mean control group	0.494	0.060	0.875	0.387	0.500
Observations	379	379	379	379	379

Note: The table shows multivariate regressions using standardized patient (SP) data. Robust standard errors are in parentheses, clustered at clinic level. Two-sided p-values in brackets. All models contain SP fixed effects and control for the 0-1 AHME treatment indicator and the malaria SP experiment (poor). The dependent variable for model 1 is a 0-1 indicator for whether the physical examination was attempted on the SP with a thermometer (temperature taken with a thermometer). The dependent variable for model 2 is a 0-1 indicator for whether the physical examination was performed on the SP with the provider's hand touching the SP's forehead (temperature taken by touch). The dependent variable for model 3 is a 0-1 indicator for whether the provider ordered any malaria diagnostic test for the SP. The dependent variable for model 4 is a 0-1 indicator for whether the provider ordered a malaria rapid diagnostic test (RDT). The dependent variable for model 5 is a 0-1 indicator for whether the provider ordered a malaria microscopy test for the SP.

b. Malaria diagnosis and treatment

	(1)	(2)	(3)	(4)	(5)	(6)
	Malaria positive	Malaria positive conditional on receiving test	Malaria RDT positive results	Malaria microscopy positive results	Dispensed/ prescribed: Artemether lumefantrine	Dispensed/ prescribed: Paracetamol
AHME treatment						
Coefficient	0.060	0.110	0.018	0.148	0.062	-0.117
Standard Error	(0.047)	(0.054)	(0.088)	(0.086)	(0.053)	(0.049)
p-value	[0.205]	[0.042]	[0.839]	[0.088]	[0.244]	[0.017]
Poor						
Coefficient	-0.013	0.043	0.257	0.001	0.015	-0.134
Standard Error	(0.059)	(0.068)	(0.109)	(0.147)	(0.063)	(0.068)
p-value	[0.831]	[0.523]	[0.021]	[0.996]	[0.816]	[0.050]
Mean control group	0.190	0.218	0.177	0.288	0.269	0.327
Observations	379	311	118	173	384	384

Note: The table shows multivariate regressions using standardized patient (SP) data. Robust standard errors are in parentheses, clustered at clinic level. Two-sided p-values in brackets. All models contain SP fixed effects and control for the 0-1 AHME treatment indicator and the malaria SP experiment (poor). The dependent variable for model 1 is a 0-1 indicator for whether the malaria SP was malaria positive unconditional on receiving any test. The dependent variable for model 2 is a 0-1 indicator for whether malaria test results were positive conditional on receiving either malaria rapid diagnostic test (RDT) or microscopy tests. The dependent variable for model 3 is a 0-1 indicator for whether the malaria RDT results were test positive given that the SP received a malaria RDT test. The dependent variable for model 4 is a 0-1 indicator for whether the malaria microscopy test results were positive given that the SP received a malaria microscopy test. The dependent variable for model 5 is a 0-1 indicator for whether the provider dispensed/prescribed artemether lumefantrine first-line treatment for malaria. The dependent variable for model 6 is a 0-1 indicator for whether the provider dispensed/prescribed paracetamol, which manages the malaria fever symptom.

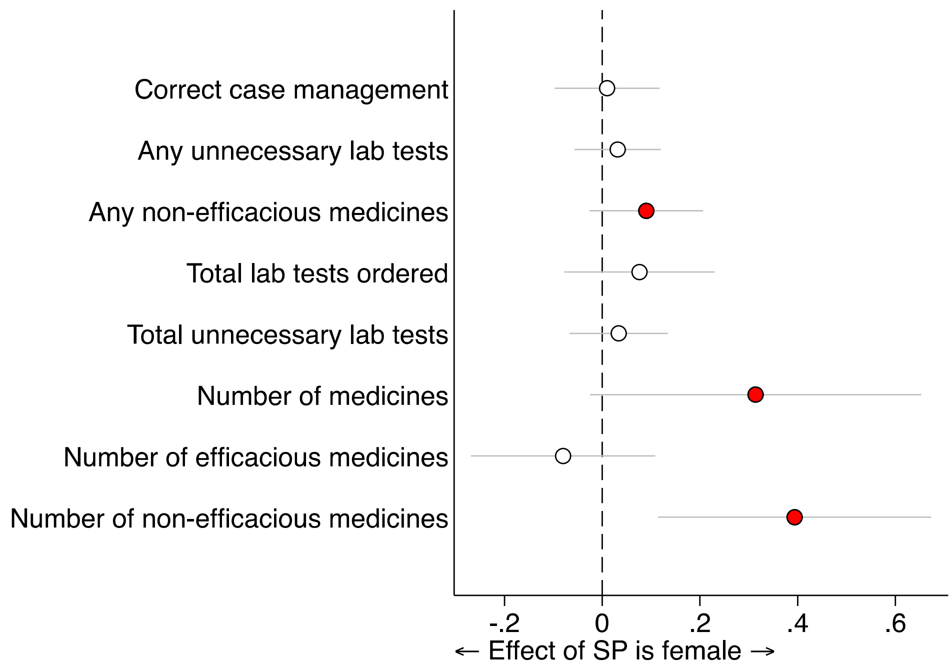
Table 6.11. Effects on Childhood Diarrhea Treatment: SP Regression Models

	(1)	(2)	(3)	(4)	(5)
	Dispensed/ prescribed: Oral rehydration salts	Dispensed/ prescribed: Zinc	Any non-efficacious medicines	Dispensed/ prescribed: Albendazole	Dispensed/ prescribed: Amoxicillin
AHME treatment					
Coefficient	0.078	-0.007	-0.104	-0.009	0.005
Standard Error	(0.052)	(0.052)	(0.048)	(0.040)	(0.031)
p-value	[0.130]	[0.888]	[0.031]	[0.831]	[0.874]
Albendazole post-demanding					
Coefficient	-0.047	-0.072	-0.015	0.137	-0.022
Standard Error	(0.061)	(0.062)	(0.058)	(0.048)	(0.037)
p-value	[0.442]	[0.243]	[0.789]	[0.004]	[0.559]
Amoxicillin post-demanding					
Coefficient	0.053	0.082	0.016	-0.156	0.025
Standard Error	(0.064)	(0.065)	(0.059)	(0.050)	(0.039)
p-value	[0.405]	[0.206]	[0.784]	[0.002]	[0.527]
Mean control group	0.352	0.420	0.723	0.205	0.091
Observations	380	380	400	380	380

Note: The table shows multivariate regressions using standardized patient (SP) data. Robust standard errors are in parentheses, clustered at clinic level. Two-sided p-values in brackets. All models contain SP fixed effects and control for the 0-1 AHME treatment indicator and the two binary 0-1 pre-post demanding phases for the childhood diarrhea SP experiments (albendazole and amoxicillin). The dependent variable in model 1 is a 0-1 indicator for whether the provider dispensed/prescribed oral rehydration salts (ORS) for the SP. The dependent variable in model 2 is a 0-1 indicator for whether the provider dispensed/prescribed zinc for the SP. The dependent variable in model 3 is a 0-1 indicator for whether the provider dispensed/prescribed any non-efficacious medicines (that is, any medicines excluding ORS and zinc) for the SP. The dependent variable in model 4 is a 0-1 indicator for whether the provider dispensed/prescribed deworming medicine albendazole, which is harmless and non-efficacious for the childhood diarrhea case. The dependent variable in model 5 is a 0-1 indicator for whether the provider dispensed/prescribed the antibiotic amoxicillin, which is harmful and non-efficacious for the childhood diarrhea case.

Figure 6.2. Effects of Gender on Process Quality: SP Regression Models for Asthma and Malaria Cases

Females were more likely to receive non-efficacious medicine but were as likely to receive lab tests—whether necessary or unnecessary—than males.



Note: The figure shows coefficients (circles) and standard error bars (whiskers) for whether the standardized patient (SP) is female from eight multivariate regression analyses on SP asthma and malaria data. Analyses contain SP case fixed effects and a 0-1 indicator for AHME treatment to control for AHME treatment effects. Dependent variables include (from top to bottom): (1) whether the case was correctly managed; (2) whether any unnecessary lab tests were ordered; (3) whether any non-efficacious medicines were prescribed/dispensed; (4) total number of lab tests ordered; (5) total number of unnecessary lab tests ordered; (6) number of medicines prescribed/dispensed; (7) number of efficacious medicines prescribed/dispensed; and (8) number of non-efficacious medicines prescribed/dispensed. Colored (red) markers indicate significant *p*-values with multiple hypothesis testing controlled at the 5% family-wise error rate using the Benjamini-Hochberg step-up procedure (Benjamini and Hochberg 1995).

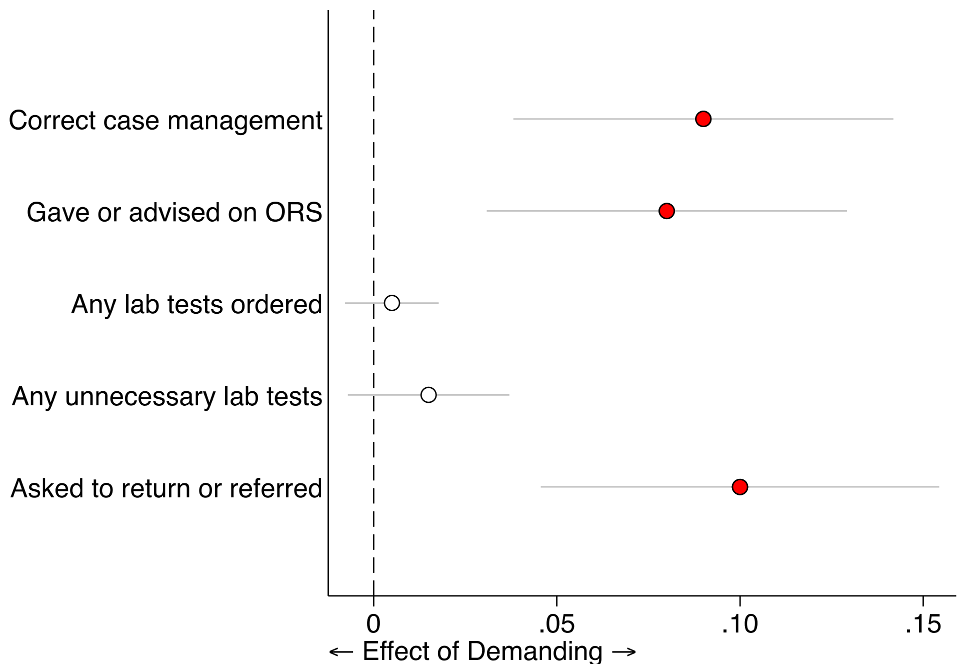
Figure 6.3. Effects of Age on Process Quality: SP Regression Models for Asthma and Malaria Cases

With each additional year of age, individuals aged 22 to 35 were more likely to receive unnecessary lab tests, but not medicines.

Note: The figure shows coefficients (circles) and standard error bars (whiskers) for SP age (22–35 years) from eight multivariate regression analyses on standardized patient (SP) asthma and malaria data. Analyses contain SP case fixed effects and a 0-1 indicator for AHME treatment to control for AHME treatment effects. Dependent variables include (from top to bottom): (1) whether the case was correctly managed; (2) whether any unnecessary lab tests were ordered; (3) whether any non-efficacious medicines were prescribed/dispensed; (4) total lab tests ordered; (5) total unnecessary lab tests ordered; (6) number of medicines prescribed/dispensed; (7) number of efficacious medicines prescribed/dispensed; and (8) number of non-efficacious medicines prescribed/dispensed. Colored (red) markers indicate significant *p*-values with multiple hypothesis testing controlled at the 5% family-wise error rate using the Benjamini-Hochberg step-up procedure (Benjamini and Hochberg 1995).

Figure 6.4. Effects of Demanding Medicines on Process Quality: SP Regression Models for Childhood Diarrhea Case

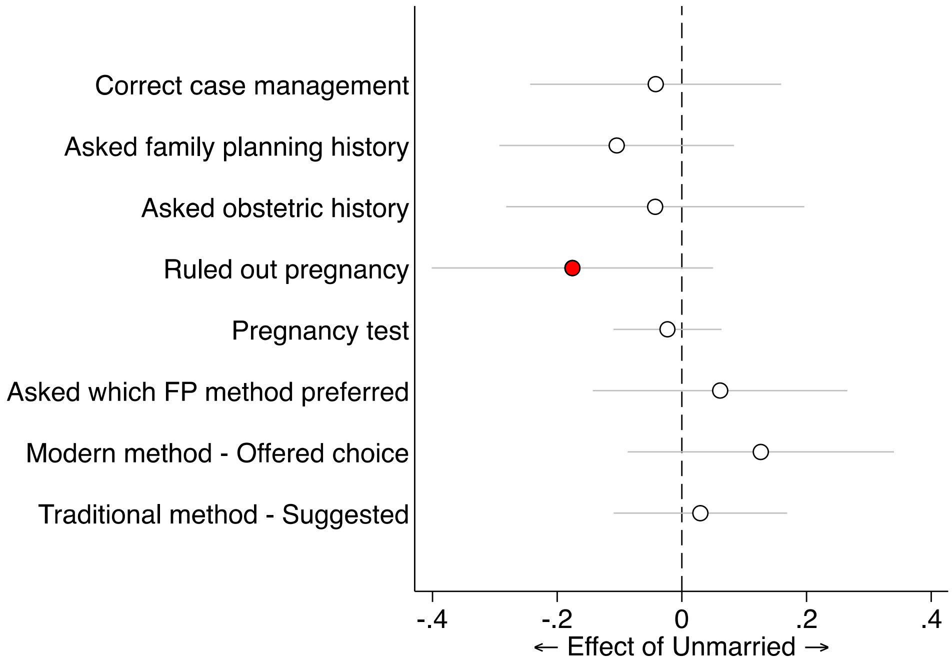
Those who demanded a non-efficacious drug received more correct care.



Note: The figure shows coefficients (circles) and standard error bars (whiskers) for the demanding standardized patient (SP) experiment (post-demanding vs. pre-demanding) from five multivariate regression analyses on SP data. Analyses contain SP fixed effects and a 0-1 indicator for AHME treatment to control for AHME treatment effects. Dependent variables include (from top to bottom): (1) whether the case was correctly managed; (2) whether the provider gave or advised on oral rehydration salts (ORS); (3) whether the provider asked the SP to return or referred the SP elsewhere; (4) whether any lab tests were ordered; and (5) whether any unnecessary lab tests were ordered. Colored (red) markers indicate significant p-values with multiple hypothesis testing controlled at the 5% family-wise error rate using the Benjamini-Hochberg step-up procedure (Benjamini and Hochberg 1995).

Figure 6.5. Effects of Being Unmarried on Process Quality: SP Regression Results for Family Planning (FP) Case

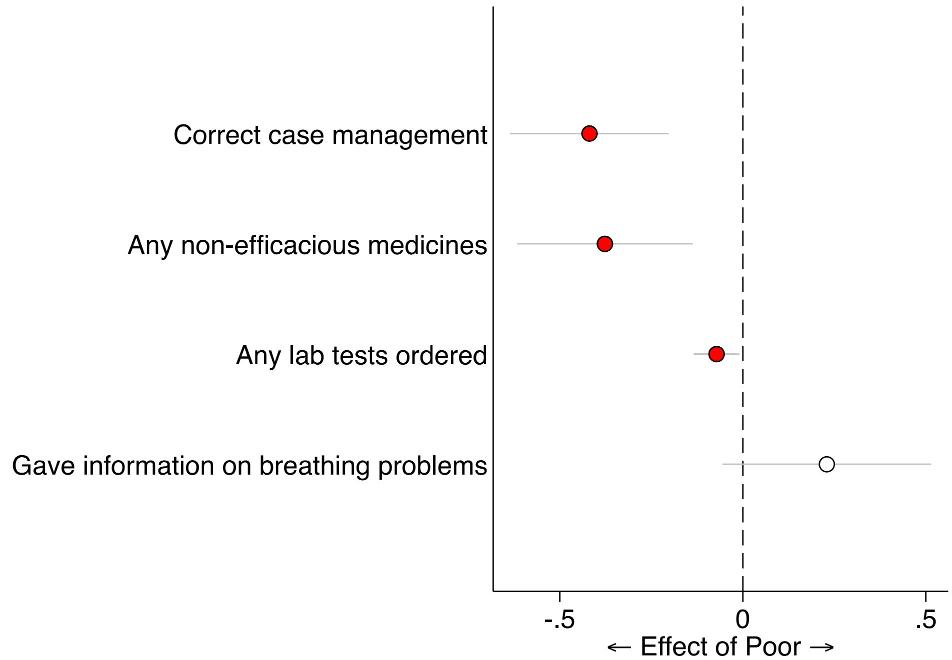
There were no observed differences between unmarried and married clients presenting for family planning services except for whether the provider ruled out pregnancy.



Note: The figure shows coefficients (circles) and standard error bars (whiskers) for the unmarried (vs. married) standardized patient (SP) experiment from eight multivariate regression analyses on SP data. Analyses contain SP fixed effects and a 0-1 indicator for AHME treatment to control for AHME treatment effects. Dependent variables include (from top to bottom): (1) whether the case was correctly managed; (2) whether the provider asked any family planning history questions; (3) whether the provider asked any obstetric history questions; (4) whether the provider ruled out pregnancy with history questions or a pregnancy test; (5) whether the provider ordered a pregnancy test; (6) whether the provider asked the client her preferred family planning (FP) method; (7) whether the provider recommended a modern method; and (8) whether the provider recommended a traditional family planning method. Colored (red) markers indicate significant *p*-values with multiple hypothesis testing controlled at the 5% family-wise error rate using the Benjamini-Hochberg step-up procedure (Benjamini and Hochberg 1995).

Figure 6.6. Effects of Being Poor on Process Quality: SP Regression Models for Asthma Case

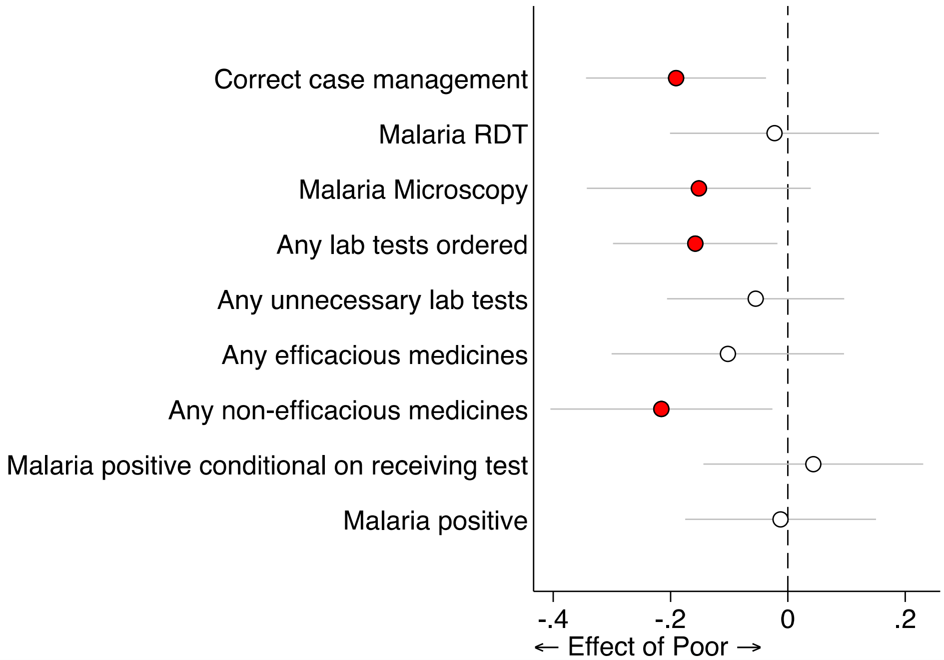
The poor got fewer instances of correct and nonessential care.



Note: The figure shows coefficients (circles) and standard error bars (whiskers) for the poor (vs. not poor) standardized patient (SP) experiment from four multivariate regression analyses on SP data. Analyses contain SP fixed effects and a 0-1 indicator for AHME treatment to control for AHME treatment effects. Dependent variables include (from top to bottom): (1) whether the case was correctly managed; (2) whether any non-efficacious medicines were prescribed/dispensed; (3) whether any lab tests (all considered unnecessary) were ordered; and (4) whether the provider gave information on breathing problems. Colored (red) markers indicate significant *p*-values with multiple hypothesis testing controlled at the 5% family-wise error rate using the Benjamini-Hochberg step-up procedure (Benjamini and Hochberg 1995).

Figure 6.7. Effects of Being Poor on Process Quality: SP Regression Models for Malaria Case

The poor were less likely to receive correct case management and microscopy and significantly less likely to receive any unnecessary medicines.



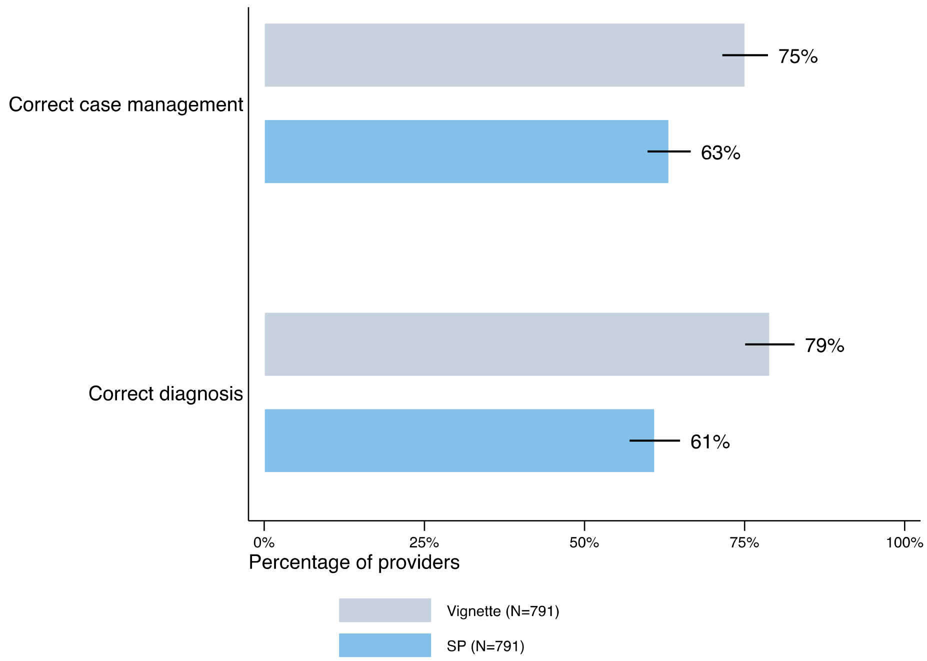
Note: The figure shows coefficients (circles) and standard error bars (whiskers) for the poor (vs. not poor) standardized patient (SP) experiment from nine multivariate regression analyses on SP data. Analyses contain SP fixed effects and a 0-1 indicator for AHME treatment to control for AHME treatment effects. Dependent variables include (from top to bottom): (1) whether the case was correctly managed; (2) whether a malaria rapid diagnostic test (RDT) was ordered; (3) whether a malaria microscopy test was ordered; (4) whether any lab tests were ordered; (5) whether any unnecessary lab tests were ordered; (6) whether any efficacious medicines were prescribed/dispensed; (7) whether any non-efficacious medicines were prescribed/dispensed; (8) whether the SP was malaria positive conditional on receiving a test; and (9) whether the SP was malaria positive unconditional on receiving a test. Colored (red) markers indicate significant *p*-values with multiple hypothesis testing controlled at the 5% family-wise error rate using the Benjamini-Hochberg step-up procedure (Benjamini and Hochberg 1995).

Table 6.12. Summary Statistics from Provider Survey Respondents

	AHME control			AHME treatment			Pooled		
	N	mean	se(mean)	N	mean	se(mean)	N	mean	se(mean)
Provider age	130	41.254	1.282	184	40.576	1.060	314	40.857	0.816
Provider is female	130	0.369	0.042	184	0.451	0.037	314	0.417	0.028
Years since training completed	130	15.046	1.241	184	15.190	1.041	314	15.131	0.796
Years spent working in health	130	15.897	1.226	184	14.485	1.027	314	15.070	0.787
Years spent working at clinic	129	8.178	0.690	184	7.216	0.553	313	7.613	0.432
Typical work week at clinic (days)	129	5.674	0.086	179	5.832	0.084	308	5.766	0.061
Locum at clinic	130	0.100	0.026	184	0.163	0.027	314	0.137	0.019
Knowledge of correct management									
Diarrhea	123	0.919	0.025	165	0.903	0.023	288	0.910	0.017
Family planning (all 4 components)	121	0.107	0.028	164	0.146	0.028	285	0.130	0.020
Family planning (at least 1 component)	122	0.934	0.023	164	0.976	0.012	286	0.958	0.012
Malaria	123	0.984	0.011	164	0.976	0.012	287	0.979	0.008

Note: Provider survey modules contain self-report measures and responses to clinical vignettes. Locum at clinic means the respondent is temporary staff at the clinic of interest. Knowledge of correct case management is a 0-1 binary measure defined in the same way as correct case management from SP surveys. For diarrhea, knowledge of correct management = 1 is defined as whether the provider mentioned giving or advising on oral rehydration salts or a referral or asking the vignette scenario to return; 0 otherwise. For family planning (all 4 components), vignette data were coded as knowledge of correct management = 1 if the provider mentioned all four of the following actions: any family planning history question, any obstetric history question, ruling out pregnancy, and asking the client her preferred family planning method; 0 otherwise. For family planning (at least 1 component), knowledge of correct management = 1 if the provider mentioned at least one of the four components; 0 otherwise. For malaria, vignette data were coded as knowledge of correct management = 1 if the provider mentioned ordering a malaria rapid diagnostic test (RDT) or a malaria microscopy test; 0 otherwise. SP refers to standardized patients.

Figure 6.8. Provider Knowledge and Practice for Correct Management and Diagnosis



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

Table 6.13. Effects on Knowledge of Correct Case Management

	(1)	(2)	(3)	(4)	(5)	(6)
	Knowledge of Correct Case Management					
	Diarrhea case		Family planning case		Malaria case	
AHME treatment						
Coefficient	-0.016	-0.006	0.039	0.041	-0.008	-0.006
Standard Error	(0.032)	(0.033)	(0.044)	(0.045)	(0.017)	(0.015)
p-value	[0.628]	[0.851]	[0.374]	[0.362]	[0.624]	[0.712]
Provider is female						
Coefficient		-0.050		0.030		-0.001
Standard Error		(0.038)		(0.049)		(0.022)
p-value		[0.188]		[0.533]		[0.978]
Provider is nurse/midwife						
Coefficient		-0.024		-0.001		-0.014
Standard Error		(0.034)		(0.051)		(0.021)
p-value		[0.485]		[0.990]		[0.510]
Provider is other staff						
Coefficient		-0.146		-0.095		-0.078
Standard Error		(0.086)		(0.043)		(0.054)
p-value		[0.092]		[0.030]		[0.149]
Mean control group	0.919	0.919	0.107	0.107	0.984	0.984
Observations	288	278	285	276	287	278

Note: The table shows multivariate regressions using provider survey data. Models 2, 4, 6 include provider gender and provider qualification covariates for Models 1, 3, 5, respectively. The constant reflects male providers who are medical doctors or clinical officers at AHME control clinics for models 2, 4, 6. Robust standard errors clustered at clinic level in parentheses; p-value in brackets. The dependent variable for all models is knowledge of correct case management, a 0-1 binary measure defined in the same way as correct case management from standardized patient (SP) surveys. For diarrhea, knowledge of correct management = 1 is defined as whether the provider mentioned giving or advising on oral rehydration salts or a referral or asking the vignette scenario to return; 0 otherwise. For family planning, vignette data were coded as knowledge of correct management = 1 if the provider mentioned all four of the following actions: any family planning history question, any obstetric history question, ruling out pregnancy, and asking the client her preferred family planning method; 0 otherwise. For malaria, vignette data were coded as knowledge of correct management = 1 if the provider mentioned ordering a malaria rapid diagnostic test (RDT) or a malaria microscopy test; 0 otherwise.

Table 7.1. AHME Effects on Client Perceptions of Amenities: SP Regression Models

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Amenities index (9-item)	Clinic is clean	Waiting time appropriate	Providers courteous and respectful	Had enough privacy	Providers spent sufficient time	Operating hours are adequate	Completely trust provider's medical treatment decision	Registration fees are reasonable	Medicines or drug fees were reasonable
AHME treatment										
Coefficient	0.008	0.007	-0.014	-0.004	0.054	0.022	0.045	-0.023	0.007	-0.027
Standard Error	(0.011)	(0.025)	(0.036)	(0.011)	(0.026)	(0.037)	(0.029)	(0.031)	(0.073)	(0.043)
p-value	[0.471]	[0.773]	[0.703]	[0.743]	[0.043]	[0.561]	[0.120]	[0.450]	[0.924]	[0.530]
Mean control group	0.792	0.890	0.810	0.966	0.822	0.498	0.852	0.782	0.731	0.724
Observations	1086	1086	1086	1086	1086	1086	1040	1076	365	805

Note: The table shows multivariate regressions using standardized patient (SP) data. Robust standard errors (in parentheses) are clustered at the clinic level. All models contain SP and case fixed effects and control for the 0-1 AHME treatment indicator and a 0-1 indicator for each SP experiment (demanding, unmarried, poor). The dependent variable for model 1 is a 9-item index ranging from 0 to 1 for the 9 SP perceptions of clinic amenities in models 2–10. The dependent variables for models 2–10 are SP perceptions of clinic amenities for: whether the clinic was clean (=1, model 2); whether the waiting time was appropriate (=1, model 3); whether providers were courteous and respectful (=1, model 4); whether clients had enough privacy (=1, model 5); whether providers spent sufficient time with the client (=1, model 6); whether operating hours were adequate (=1, model 7); whether the SP completely trusted the provider's treatment decision (=1, model 8); whether the SP thought registration fees were reasonable, if applicable (=1, model 9); and whether the SP thought medicine or drug fees were reasonable, if applicable (=1, model 10).

Table 7.2. AHME Effects on Client Perceptions of Amenities: Client Exit Interview Regression Models

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Amenities index (9-item)	Clinic is clean	Waiting time appropriate	Providers courteous and respectful	Had enough privacy	Providers spent sufficient time	Operating hours are adequate	Completely trust provider's medical treatment decision	Registration fees are reasonable	Medicines or drug fees were reasonable
AHME treatment										
Coefficient	0.006	0.027	0.005	0.003	0.010	-0.033	0.001	0.032	-0.042	-0.022
Standard Error	(0.010)	(0.017)	(0.014)	(0.008)	(0.019)	(0.026)	(0.020)	(0.023)	(0.048)	(0.030)
p-value	[0.528]	[0.113]	[0.722]	[0.732]	[0.590]	[0.204]	[0.955]	[0.163]	[0.378]	[0.455]
Mean control group	0.858	0.922	0.929	0.974	0.894	0.544	0.884	0.888	0.830	0.834
Observations	1539	1532	1537	1535	1534	1531	1508	1528	356	828

Note: The table shows bivariate regressions using client exit interview data. Robust standard errors (in parentheses) are clustered at the clinic level. All models contain the binary 0-1 AHME treatment indicator. The dependent variable for model 1 is a 9-item index ranging from 0 to 1 for the 9 client perceptions of clinic amenities in models 2–10. The dependent variables for models 2–10 are client perceptions of clinic amenities for: whether the clinic was clean (=1, model 2); whether the waiting time was appropriate (=1, model 3); whether providers were courteous and respectful (=1, model 4); whether clients had enough privacy (=1, model 5); whether providers spent sufficient time with the client (=1, model 6); whether operating hours were adequate (=1, model 7); whether the client completely trusted the provider's treatment decision (=1, model 8); whether the client thought registration fees were reasonable, if applicable (=1, model 9); and whether the client thought medicine or drug fees were reasonable, if applicable (=1, model 10).

Table 7.3. AHME Effects on SP Satisfaction Rating (1, low, to 10, high) (z-score): SP Regression Models

	(1) Diarrhea	(2) Family planning	(3) Asthma	(4) Malaria	(5) Pooled
AHME treatment					
Coefficient	0.058	0.141	-0.033	-0.068	0.018
Standard Error	(0.101)	(0.145)	(0.187)	(0.116)	(0.091)
p-value	[0.569]	[0.333]	[0.862]	[0.555]	[0.844]
Mean control group	0.000	0.000	0.000	0.000	0.000
Observations	378	197	132	379	1086

Note: The table shows multivariate regressions using standardized patient (SP) data. Standard errors are in parentheses and p-values are in brackets. All models contain SP fixed effects and control for the 0-1 AHME treatment indicator and a 0-1 indicator for each SP experiment (demanding, unmarried, poor). Model 5 contains case fixed effects. The dependent variable in models 1–5 is an SP satisfaction rating from 1 (least) to 10 (most) the SP gave to the provider, and then standardized to the mean of the AHME control group.

Table 7.4. AHME Effects on Provider Did a Good Job Explaining (0-1): SP and Client Exit Interview Regression Models

	(1)	(2)	(3)	(4)	(5)	(6)
	Standardized Patients					***Exit Interviews***
	Diarrhea	Family planning	Asthma	Malaria	Pooled	Pooled
AHME treatment						
Coefficient	-0.063	0.150	-0.051	0.015	0.002	0.002
Standard Error	(0.041)	(0.070)	(0.077)	(0.048)	(0.034)	(0.012)
p-value	[0.130]	[0.034]	[0.506]	[0.752]	[0.954]	[0.841]
Mean control group	0.826	0.567	0.776	0.649	0.714	0.950
Observations	378	197	132	379	1086	1530

Note: The table shows multivariate regressions using standardized patient (SP) data (models 1, 2, 3, 4, 5) and bivariate regressions using client exit interview data (model 6). Standard errors are in parentheses and p-values are in brackets. Models 4, 5, and 6 contain robust standard errors (in parentheses) clustered at the clinic level. All models contain the binary 0-1 AHME treatment indicator for the clinic visited by the standardized patients or the actual clients. Models 1–5 contain SP fixed effects and a 0-1 indicator for each SP experiment (demanding, unmarried, poor). Model 5 contains case fixed effects. The dependent variable for models 1–6 is a 0-1 indicator for whether the SP or actual client found the provider did a good job explaining.

Table 7.5. AHME Effects on Household Perceptions of the Index Clinic: Pooled Regression Models

	(1)	(2)
	Index clinic quality rating 1 (low)–10 (high) (z-score)	Index clinic ranked as high quality for visit (z-score)
AHME treatment		
Coefficient	0.039	0.050
Standard Error	(0.054)	(0.048)
p-value	[0.471]	[0.302]
Visit: Well baby		
Coefficient	-0.007	0.045
Standard Error	(0.038)	(0.049)
p-value	[0.855]	[0.356]
Visit: Child curative		
Coefficient	-0.249	-0.075
Standard Error	(0.038)	(0.038)
p-value	[0.000]	[0.052]
Mean control group	0.000	0.000
Observations	2637	2637
Households	1139	1139
Clinics	192	192

Note: The table shows multivariate regressions using household survey data. Robust standard errors are in parentheses, clustered at level of index clinic. Two-sided p-values in brackets. Clients were asked to rank the quality of services at the index clinic compared to other clinics they had visited. The dependent variable in model 1 is the numerical rating of the index clinic, transformed into a z-score with mean control group zero. The dependent variable in model 2 is an indicator for if the index clinic was ranked in the 1st–50th percentile among the respondent's choice set, where a higher quality clinic is ranked lower. Households gave their opinion on three types of visits: child curative, child preventative, and prenatal care. These models include responses about all three types of visits. We control for the type of visit, where prenatal care is the reference category.

Table 7.6. AHME Effects on Client Experience: SP and Client Exit Interview Regression Models

	(1)	(2)	(3)	(4)	(5)
	Standardized Patients			***Client Exit Interviews***	
	Number of clients waiting (z-score)	Minutes spent with provider (z-score)	Minutes spent at clinic (z-score)	Number of clients waiting (z-score)	Minutes spent waiting (z-score)
AHME treatment					
Coefficient	-0.111	0.014	0.034	0.055	0.038
Standard Error	(0.098)	(0.075)	(0.084)	(0.107)	(0.079)
p-value	[0.260]	[0.848]	[0.683]	[0.609]	[0.632]
Mean control group	0.000	0.000	0.000	0.000	0.000
Observations	1195	1086	1195	1539	1539

Note: The table shows multivariate regression results from standardized patients (SP) data (models 1, 2, 3) and bivariate regression results from client exit interview data (models 4 and 5). All models contain the binary 0-1 AHME treatment indicator and robust standard errors in parentheses, clustered at the clinic level. Two-sided p-values in brackets. Models 1, 2, and 3 contain SP and case fixed effects and control for the 0-1 AHME treatment indicator and a 0-1 indicator for each SP experiment (demanding, unmarried, poor). Dependent variables in models 1–5 are z-score measures standardized to the AHME mean control group. Number of clients waiting in model 1 refer to the number of individuals in the waiting room upon SP arrival at the clinic. Minutes spent with the provider is the total time spent by the SP with the main provider(s) (model 2). Minutes spent at clinic is the time from when the SP entered and exited the clinic (model 3). Number of clients waiting in model 4 refers to the number of individuals in the waiting room when the interviewed client arrived. Minutes spent waiting in model 5 is the total time spent waiting at the clinic for the client.

Table 8.1 Descriptive Statistics from Modified Dictator Game

	N	Mean	Std. Dev.	Min	Max
a) Provider Survey Respondents					
Client Budget Share - All Decisions	302	0.54	0.20	0.00	0.96
Client Budget Share - Real Client Decisions	302	0.48	0.22	0.00	1.00
Client Budget Share - Not Poor Client Decisions	302	0.51	0.20	0.00	0.94
Client Budget Share - Poor Client Decisions	301	0.62	0.25	0.00	1.00
Client Budget Share - Malaria Poor Decisions	301	0.62	0.25	0.00	1.00
Client Budget Share - Malaria Standard Decisions	301	0.49	0.23	0.00	1.00
Client Budget Share Difference (Poor - Standard)	301	0.13	0.20	-0.46	0.83
Afriat CCEI Violation at 1.00	300	0.17	0.38	0.00	1.00
Afriat CCEI Violation at 0.95	300	0.00	0.00	0.00	0.00
Afriat CCEI Violation at 0.80	300	0.00	0.00	0.00	0.00
Alpha	300	0.65	0.26	0.00	1.00
SE(Alpha)	300	0.09	0.08	0.00	0.42
Rho	300	-2.61	5.69	-20.00	0.97
SE(Rho)	300	3.36	2.87	0.00	9.21
Sigma	300	-0.96	2.24	-35.32	-0.05
b) SP Visits Matched to Provider Survey Respondents					
	N	Mean	Std. Dev.	Min	Max
Client Budget Share - All Decisions	859	0.52	0.21	0.00	0.96
Client Budget Share - Real Client Decisions	859	0.47	0.23	0.00	1.00
Client Budget Share - Not Poor Client Decisions	859	0.49	0.21	0.00	0.94
Client Budget Share - Poor Client Decisions	856	0.61	0.25	0.00	1.00
Client Budget Share - Malaria Poor Decisions	856	0.61	0.25	0.00	1.00
Client Budget Share - Malaria Standard Decisions	856	0.48	0.24	0.00	1.00
Client Budget Share Difference (Poor - Standard)	856	0.13	0.20	-0.46	0.83
Afriat CCEI Violation at 1.00	850	0.17	0.37	0.00	1.00
Afriat CCEI Violation at 0.95	850	0.00	0.00	0.00	0.00
Afriat CCEI Violation at 0.80	850	0.00	0.00	0.00	0.00
Alpha	850	0.67	0.24	0.00	1.00
SE(Alpha)	850	0.09	0.08	0.00	0.42
Rho	850	-2.33	5.41	-20.00	0.97
SE(Rho)	850	3.12	2.84	0.00	9.21
Sigma	850	-1.07	2.96	-35.32	-0.05

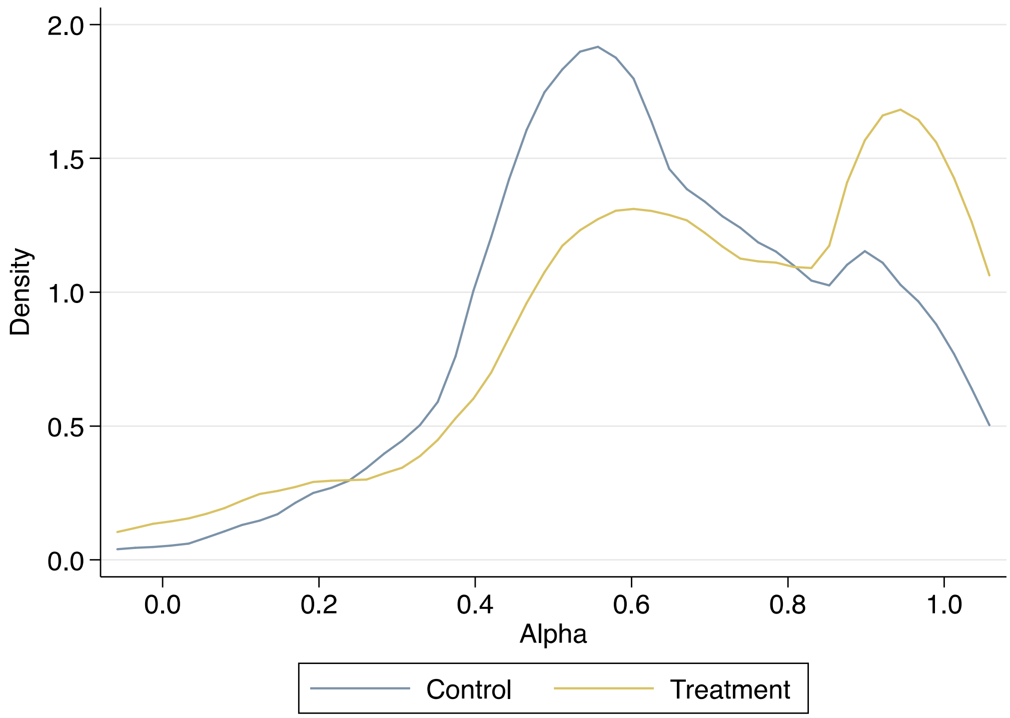
Note: Top panel a) shows summary statistics from respondents of the provider survey; bottom panel b) shows summary statistics from the respondents of the provider survey matched to providers seen in SP visits, where the same interviewed provider could have successfully seen and been identified for one or more SP visits. Client Budget Share is the amount client was allocated divided by the sum of the amount provider kept and amount provider allocated to client. Afriat's Critical Cost Efficiency Index (CCEI) Violation at 1.00, 0.95, 0.80 are three binary (0-1) indicators for whether the provider corresponding to the observation violates Afriat's theorem for consistency for fitting a standard utility function at the 1.00, 0.95, or 0.80 Afriat's (1972) CCEI threshold, respectively. Following the literature, we call individuals 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. Sigma = 1 / (rho - 1) is the constant elasticity of substitution where rho <= 1 is the willingness to trade off payments to self and other in response to price changes. When rho > 0 (rho < 0), an individual's social preferences are weighted towards efficiency (equality). Observations are not consistent within panels because we are only able to determine Afriat's CCEI threshold violations and social preference parameters (alpha, rho, sigma) for provider survey respondents who conducted all decisions in the modified dictator game.

Table 8.2 AHME Effects on Client Budget Share Allocated in a Modified Dictator Game

		(1)	(2)
		Client Budget Share	
AHME treatment			
	Coefficient	-0.090	-0.072
	Standard Error	0.000	0.000
	p-value	[0.000]	[0.000]
Client Scenario is Poor			
	Coefficient	0.141	0.141
	Standard Error	(0.012)	(0.012)
	p-value	[0.000]	[0.000]
Game Multiplier			
	Coefficient	0.041	0.041
	Standard Error	(0.003)	(0.003)
	p-value	[0.000]	[0.000]
Childhood Diarrhea Client Recipient			
	Coefficient	0.070	0.070
	Standard Error	(0.012)	(0.012)
	p-value	[0.000]	[0.000]
Malaria Client Recipient			
	Coefficient	0.009	0.009
	Standard Error	(0.010)	(0.010)
	p-value	[0.365]	[0.365]
Family Planning Client Recipient			
	Coefficient	-0.010	-0.010
	Standard Error	(0.011)	(0.011)
	p-value	[0.365]	[0.365]
Respondent is in-charge			
	Coefficient		-0.018
	Standard Error		0.000
	p-value		[0.000]
Mean control group		0.563	0.563
Observations		5120	5120

Note: Table shows multivariate OLS regressions using provider survey data from a modified dictator game where each observation is at the provider-decision level. The constant reflects providers allocating an endowment to real clients at AHME control clinics. Robust standard errors clustered at clinic level in parentheses; p-value in brackets. All models include covariates for AHME treatment (0-1) and whether the recipient client scenario was poor (0-1) and control for recipient client scenario (fictitious childhood diarrhea client, fictitious malaria client, fictitious family planning client) and provider fixed effects. Model 2 includes a binary control variable for whether the respondent was an in-charge (n = 210 of total 403 respondents were in-charges; in-charge status of 101 of 403 respondents is unknown and coded as missing). Client budget share refers to the amount the respondent allocated to the client divided by the sum of the amount the respondent kept and the amount allocated to the client.

Figure 8.1 Conditional Density of Social Preference Parameter: Alpha



Note: Figure shows kernel density plots of SP data and provider survey data from a modified dictator game where each observation is one SP-provider interaction. We elicited the social preference parameter alpha from a modified dictator game in our provider survey. 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.

Table 8.3 AHME Effects on Correct Case Management: Least Altruistic

	(1)	(2)	(3)	(4)
	Correct case management			
AHME treatment				
Coefficient	-0.066	-0.050	-0.048	-0.044
Standard Error	(0.033)	(0.041)	(0.042)	(0.042)
p-value	[0.048]	[0.226]	[0.253]	[0.294]
Least altruistic (50%)				
Coefficient		0.019		
Standard Error		(0.040)		
p-value		[0.630]		
Least altruistic (25%)				
Coefficient			0.001	
Standard Error			(0.044)	
p-value			[0.974]	
Least altruistic (20%)				
Coefficient				-0.030
Standard Error				(0.048)
p-value				[0.526]
Mean control group	0.690	0.690	0.690	0.690
Observations	1086	802	802	802

Note: Table shows multivariate OLS regressions using SP data and provider survey data from a modified dictator game where each observation is one SP-provider interaction. Robust standard errors clustered at clinic level in parentheses; two-sided *p*-value in brackets. All models include covariates for AHME treatment (0 if clinic was in AHME control and 1 if clinic was in AHME treatment), each SP experiment (demanding, unmarried, poor), and control for SP and case fixed effects. The outcome for models 1-4 is correct case management a binary indicator for whether the SP visit was correctly managed (0 if no, 1 if yes). Model 1 shows the effect of AHME. Models 2, 3, and 4 is the same as Model 1 but includes an indicator for whether the provider seen by the SP is 50%, 25%, 20% least altruistic, respectively based on our continuous altruism parameter. Altruistic indicators are the bottom percentage of negative alpha: the bottom 50% is the 50% least altruistic (i.e., most self-interested), ..., and the bottom 20% is the 20% least altruistic. 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.

Table 8.4 Heterogeneity of AHME Effects on Correct Case Management: Least Altruistic

	(1)	(2)	(3)	(4)	(5)
			Correct case management		
AHME treatment					
Coefficient	-0.066	-0.073	0.024	-0.010	-0.011
Standard Error	(0.033)	(0.035)	(0.061)	(0.050)	(0.048)
p-value	[0.048]	[0.040]	[0.690]	[0.850]	[0.822]
AHME treatment * Least Altruistic (50%)					
Coefficient			-0.159		
Standard Error			(0.088)		
p-value			[0.074]		
AHME treatment * Least Altruistic (25%)					
Coefficient				-0.189	
Standard Error				(0.105)	
p-value				[0.073]	
AHME treatment * Least Altruistic (20%)					
Coefficient					-0.227
Standard Error					(0.121)
p-value					[0.063]
Mean control group	0.690	0.679	0.679	0.679	0.679
Observations	1086	958	674	674	674

Note: Table shows multivariate OLS regressions using SP data and provider survey data from a modified dictator game where each observation is one SP-provider interaction. Robust standard errors clustered at clinic level in parentheses; two-sided *p*-value in brackets. All models include covariates for AHME treatment (0 if clinic was in AHME control and 1 if clinic was in AHME treatment), each SP experiment (demanding, unmarried, poor), and control for SP and case fixed effects. The outcome correct case management is a binary indicator for whether the SP visit was correctly managed. Model 1 includes the full sample with providers correctly identified and replacements for those providers who were not identified (same-clinic and same-service). Models 2-5 restrict the sample to SP visits conducted by correctly identified and matched providers from the provider survey. Model 1 does not include altruism parameters. Model 2 does not include altruism parameters and restricts the sample as described. Models 3, 4, and 5 include indicators for whether the provider seen by the SP is 50%, 25%, 20% least altruistic, respectively based on our continuous altruism parameter: Models {3,4,5} includes a binary variable to indicate whether the provider falls in the {50%, 25%, 20%} least altruistic group and interactions between AHME treatment (0-1) and the binary variable for whether the provider falls in the {50%, 25%, 20%} least altruistic group. Altruistic indicators are the bottom percentage of negative alpha: the bottom 50% is the 50% least altruistic (i.e., most self-interested), ..., and the bottom 20% is the 20% least altruistic. 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.

Table 8.5 Heterogeneity of AHME Effects on Prices: Least Altruistic

	(1)	(2)	(3)	(4)	(5)
			Price (KSH) (arcsinh)		
AHME treatment					
Coefficient	0.103	0.088	0.053	0.091	0.104
Standard Error	(0.177)	(0.212)	(0.339)	(0.273)	(0.263)
p-value	[0.560]	[0.677]	[0.876]	[0.738]	[0.694]
AHME treatment * Least Altruistic (50%)					
Coefficient			0.712		
Standard Error			(0.563)		
p-value			[0.207]		
AHME treatment * Least Altruistic (25%)					
Coefficient				1.509	
Standard Error				(0.737)	
p-value				[0.042]	
AHME treatment * Least Altruistic (20%)					
Coefficient					2.233
Standard Error					(0.822)
p-value					[0.007]
Semi-elasticity of AHME					0.020
p-value AHME					0.693
Semi-elasticity of Least Altruistic (20%) in AHME					1.212
p-value Least Altruistic (20%) in AHME					0.003
Mean control group	5.367	5.258	5.258	5.258	5.258
Observations	684	584	401	401	401

Note: Table shows multivariate OLS regressions using SP data and provider survey data from a modified dictator game where each observation is one SP-provider interaction. Robust standard errors clustered at clinic level in parentheses; p-value in brackets. These data exclude childhood diarrhea pre-demanding (since price is captured once post-demanding) and asthma poor and malaria poor (since price was capped by design at 300 KSH) SP visits. All models include covariates each SP experiment (unmarried), and control for SP and case fixed effects. The outcome is the total amount the SP paid (in Kenyan Shillings, KSH) transformed to hyperbolic arcsine. Model 1 includes the full sample with providers correctly identified and replacements for those providers who were not identified (same-clinic and same-service). Models 2-5 restrict the sample to SP visits conducted by correctly identified and matched providers from the provider survey. Model 1 does not include altruism parameters. Model 2 does not include altruism parameters and restricts the sample as described. Models 3, 4, and 5 include indicators for whether the provider seen by the SP is 50%, 25%, 20% least altruistic, respectively based on our continuous altruism parameter: Models {3,4,5} includes a binary variable to indicate whether the provider falls in the {50%, 25%, 20%} least altruistic group and interactions between AHME treatment (0-1) and the binary variable for whether the provider falls in the {50%, 25%, 20%} least altruistic group. Altruistic indicators are the bottom percentage of negative alpha: the bottom 50% is the 50% least altruistic (i.e., most self-interested), ..., and the bottom 20% is the 20% least altruistic. 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.

Appendices

Supplemental appendices available for this dissertation include:

- Chapter 5 Supplement for Structures: Tables and Figures
- Chapter 6 Supplement for Processes: Text
  - Standardized Patients: Supplementary Methodology
  - Standardized Patients: Ethical Considerations
  - Standardized Patients: Request for Waiver of Informed Consent
  - Standardized Patients: Childhood Diarrhea SP Case (Example)
  - Provider Survey: Vignettes
- Chapter 6 Supplement for Processes: Tables and Figures
- Chapter 7 Supplement for Health Care Outcomes: Tables and Figures
- Chapter 8 Supplement for Provider Preferences: Altruism vs. Self-interest

These are available for download at: `https://github.com/kwantify/dissertation`.