

**Essays on Health, Human Rights, and  
Social Development**

by

Joshua Louis Greenberg

A dissertation presented in partial fulfillment  
of the requirements for the degree of  
Doctor of Philosophy  
(Economics)  
in The University of Michigan  
2021

Doctoral Committee:

Professor C. Hoyt Bleakley Jr., Chair  
Professor John Bound  
Lecturer Fred Muhumuza, Makerere University  
Professor Edward Norton  
Professor Dean Yang

Joshua Louis Greenberg

jlgr@umich.edu

ORCID iD: 0000-0002-2928-8111

© Joshua Louis Greenberg 2021

## **DEDICATION**

To all those throughout the world who fight for truth, justice, equality, freedom, and peace, and who believe that togetherness is our greatest asset in creating change.

To all those in Uganda who have opened my eyes to the depths of humanity and to the complexities of social systems.

To all those in my life who have lifted, inspired, supported, cultivated, pushed, embraced, welcomed, tolerated, and withstood me so that I could reach this moment.

## ACKNOWLEDGEMENTS

Writing this dissertation would have been impossible without the generous support of numerous colleagues, mentors, and friends, all of whom I cannot possibly list here. Let me, however, touch on as many of them as possible. First, I am eternally grateful to the members of my dissertation committee, who have patiently worked with me over the course of several years to help create this final product. I especially thank the chair, Hoyt Bleakley, for his boundless insight, his unwavering commitment to my professional and intellectual growth, and his reliable, witty humor. Hoyt has been the best I could have ever asked for from a main advisor — giving me all of the freedom I needed to explore and innovate, while simultaneously always pushing me to go further. My ineffable gratitude also goes to John Bound, who played an instrumental role in accepting me into the MD/PhD program and who has been one of the strongest advocates of my work over the course of my time at the University of Michigan. Were it not for John’s remarkable character, I would not have even had a chance to write this dissertation. With much of my research situated in Uganda, Fred Muhumuza has gone to tremendous lengths to offer me an intellectual home in the country. Most of all, he has contributed a profound combination of cultural and scientific insight toward my research and facilitated critical relationships to assist in my work. In addition, given my interdisciplinary interests, Edward Norton has provided an essential public health perspective on my research. I can always count on Edward to make key observations that have not crossed anyone else’s mind, while also fostering a most enjoyable work environment. Finally, Dean Yang has been a true role model for field-based research in development economics and has been a constant source of inspiration for my work. Both his enthusiasm and intellect have kept me going on multiple occasions.

At the University of Michigan Medical School, where I am jointly enrolled, I thank Ron Koenig, the former director of the Medical Scientist Training Program (MSTP). Together

with John Bound, Ron has been one of the greatest advocates of my work since day one, always ready to muster on my behalf any resources that he could. In addition, I thank Joe Kolars, the Senior Associate Dean for Education and Global Initiatives, who so graciously welcomed me to the University. Joe has been an outstanding mentor and collaborator in global health work, going to unnecessary lengths to support my dissertation research. I also thank Kathy Collins, the new director of the MSTP, whom I have had the privilege of getting to know this past year and who has proven to be a remarkable source of support in filling Ron's shoes.

Stretching back further, I am grateful to four undergraduate mentors: Alessandro Tarozzi, Frank Sloan, Subhrendu Pattanayak, and Alex Cho. Alessandro weathered the endless demands that I made on his time, all while cultivating my interest in the elegant and powerful discipline that is economics. Frank was the first individual — and one of only a few — to embrace my desire to pursue advanced degrees in both medicine and economics. Moreover, some of the key lessons that Frank taught me planted the intellectual seeds that, over many years, germinated into Chapters 1 and 2 of this dissertation. Subhrendu gave me a model of economics in the real world, demonstrating the versatile applications of the discipline in improving people's lives. Alex, an entrepreneurial physician, nurtured my interest in global health and made indispensable, selfless contributions to my pursuits, including the groundwork that made Chapter 3 of this dissertation possible. Alongside all of the foregoing individuals from my undergraduate and graduate institutions, I thank my longtime friend and mentor Peter Mugenyi, the pioneering HIV/AIDS leader and scholar. Peter has served as a sounding board for all of my research and has always offered his wisdom and support in furtherance of my work.

Pulling off intensive field projects, as in this dissertation, requires countless helping hands and moving parts to come together. At each stage of project development, the activities often hinge on the goodwill of others. Without this goodwill, nothing would come to fruition. Thus, I am forever grateful for this generosity and can only hope that the impacts of my research match the extent of the support that has been given to me. For Chapters 1 and 2 specifically, many public servants in the Ugandan government have made extraordinary contributions to make this work possible. An enormous thank you goes to the staff of the

Department of Monitoring, Evaluation, and Inspection at the Office of the Prime Minister, including Commissioner Timothy Lubanga, Assistant Commissioner Gonzaga Mayanja, and Joseph Muserero. All three provided essential input for this work. Most of all, without the selflessness, dedication, and leadership of Timothy, the work for Chapters 1 and 2 would have had little hope of advancing. I also thank Geoffrey Mugumya of the Ministry of Local Government for initially recognizing the promise of this work and introducing me to the Office of the Prime Minister.

Several colleagues at the Uganda Bureau of Statistics (UBOS) have offered critical research support for the first two chapters. I thank Bernard Muhwezi, for always connecting me to whatever resources I needed, whenever I needed them, and for being an ever reliable source of encouragement and humor; Wilson Nyegenye, for providing background census data; Vincent Ssenono, for his guidance on sampling; and Byron Twesigye, for his advice on pre-survey household enumeration. I also thank my longtime colleague, Margaret Ayebare, for helping to facilitate my relationship with UBOS and for always energizing me to work harder. At the Electoral Commission, I thank Charles Lubwama for providing 2016 local elections data. At the Ministry of Health, I thank Samson Olum for so willingly assisting with health-related government resources. Equally important, I owe a great debt to multiple colleagues at the University of Michigan Institute for Social Research (ISR) for their research support: Stephanie Chardoul, for her coordination of the ISR team; Lisa Holland, for her review of questionnaires and her remarkable insight into question formulation; Margaret Levenstein, for her guidance on data management and grant-writing; and Raphael Nashimura, for his flowing generosity in advising on survey methodology.

Three more colleagues played essential roles in the early formulation of Chapters 1 and 2. First, I thank Pia Raffler, of Harvard, for all of the time she spent talking through ideas and plans and for so generously sharing research instruments and implementation tools from her related work. Without Pia's contributions, I am certain that the project would have faltered. Second, Nasul Kabunga, of the International Food Policy Research Institute, offered significant insight from his previous research and entertained countless questions from me on how best to go about implementing the activities. Third, Ibrahim Kasirye, of the Economic Policy Research Centre at Makerere University, offered tremendous advice on

the health facility quality assessment and provided a similar tool from his related work.

For Chapter 3, I extend gratitude to Katherine Donato, of Boston Consulting Group, who invested significant time in the development and management of the data collection. I also thank Joseph Ngonzi, of Mbarara University of Science and Technology, for his indispensable leadership role in the management of the project. Chapter 4 would not have been possible without Hoyt Bleakley, with whom the work is co-authored. It is thanks to Hoyt's brilliance, spanning theory and empirics, that this chapter came about. We also thank Mike Mueller-Smith for his advice on our empirical approach.

Cutting across the dissertation, three of the chapters are based on projects in Uganda, while a fourth chapter has been inspired by observations I have made over the course of all of the time I have spent working on community development in the country. All of the activities for this work have been implemented by Progressive Health Partnership (PHP), an organization that I co-founded.<sup>1</sup> Words cannot describe my gratitude for everything that the PHP team has done. It has been the privilege of a lifetime to work together. I express ineffable thanks to Jordan Bateisibwa, the director of PHP in Uganda, for his overall management of the research, programmatic, and administrative components of the work that went into Chapters 1, 2, and 3. Jordan's leadership skills, insight, enthusiasm, and passion have consistently elevated our activities to the next level. For the first two chapters, none of the data collection would have been possible without an extraordinarily hardworking field research team, including a group of interviewers who repeatedly opened my eyes to new ways of thinking about data collection. I especially thank Sam Twesigye, the field manager, whose indefatigability, dedication, and innovation — even in the face of an overstretched project — blew me away time and time again and make him a model for all of us to live up to. In addition, Justine Kyasimire played a key role in the qualitative field research involved in the initial conceptualization of the activities. On the program implementation side, Joram Amany, Ruth Kaziga, and Alex Niwagaba worked a grueling schedule to carry out field activities, always striving to uphold the highest standards of quality. For Chapter 3, Joram Amany, Diana Asimwe, Ephratah Nahabwe, and Alex Niwagaba played critical

---

<sup>1</sup>I currently serve as the Chief Executive Officer and as a board member of PHP. My relationship with the organization is unpaid.

roles in the data collection and program implementation, from the early foundations of the project through its completion several years later. For all of PHP's activities, I also owe profound appreciation to Christine Convery for her diligent financial management and her administrative acumen. Without all of Christine's behind-the-scenes work, our project teams would have been lost.

Equally as important as the implementation teams, none of the work would have been possible without the local community members who have the grace and courage to welcome us into their lives. These community members have taught me more about global health and development than any academic course and have provided all of the core ideas that have underlain our work in Uganda.

Financially, my dissertation research would not have been possible without generous funding from multiple sources:

- Abdul Jameel Latif Poverty Action Lab (J-PAL): Governance Initiative Award (grant number GR-0947)
- Eunice Kennedy Shriver National Institute of Child Health and Human Development of the National Institutes of Health: Training Fellowship at Population Studies Center, University of Michigan Institute for Social Research (3 years, award number T32HD007339)
- Nancy Allison Perkins Foundation: Program Grant
- National Institute of General Medical Sciences of the National Institutes of Health: Medical Scientist Training Program Fellowship, University of Michigan Medical School (award number T32GM007863)
- National Science Foundation: Doctoral Dissertation Research Improvement Grant (grant number 1824290)
- U.S. Department of Education: Fulbright-Hays Doctoral Dissertation Research Abroad Program (grant number P022A180037)
- University of Michigan African Studies Center and International Institute: Student Fellowship (three awards)



- University of Michigan Department of Economics: Jordan and Kim Dickstein African Initiative Fund Research Award
- University of Michigan Department of Economics: Summer Research Apprenticeship (four awards)
- University of Michigan Medical School Global REACH Program: Global Health Project Grant
- University of Michigan Medical School Global REACH Program: Global Health Travel Grant
- University of Michigan Population Studies Center: Weinberg Research Prize
- University of Michigan Rackham Graduate School: Rackham International Research Award

The content of the dissertation is solely my responsibility and does not necessarily represent the official views of any of the funders.

I especially thank Nancy and Bob Perkins and Lindsay Vail for their outstanding financial and moral support toward PHP's work in Uganda over the course of many years. Their extraordinary commitment to the organization laid the foundation for all of my dissertation research. For life-saving administrative support in managing the various sources of funding, I thank Jake Crawford, Khristine Henson-Jones, and Julie Esch of the University of Michigan Department of Economics Business Office and Beth Dutridge-Corp of the University of Michigan International Institute. At Makerere University, I thank Martha Muwanguzi of the International Affairs Office for her tremendous assistance in facilitating my local academic affiliation. In addition, I am grateful to Dorothy Ngalombi, of the U.S. Embassy in Kampala, who, in helping to manage two different Fulbright awards over several years, has gone to great lengths to support my work.

Lastly, and certainly not least, no set of acknowledgements would be complete without thanking my parents, without whom I would have had no hope of advancing to a PhD. It is because of their perseverance, combined with the freedom that they have always afforded me to do what I love, that I am here today.

## PREFACE

This dissertation presents a diverse set of research related to three broad topics: health, human rights, and social development. The first area, health, is the most concrete subject matter of my work, particularly in light of my dual course of study in medicine and economics. In the dissertation, I address health from a number of perspectives. Chapters 1 and 2 are oriented around a program aimed at fostering greater local political accountability in Uganda to promote the provision of higher-quality health services. While these chapters primarily address supply-side factors in healthcare delivery, Chapter 3 examines the demand-side, looking at the impact of a community health worker intervention on maternal health behaviors and outcomes in Uganda. Then, in contrast to the focus on health services in the first three chapters, Chapter 4 draws on the role of health in human development and human capital formation, and is motivated by the notion that lifespan often affects the economic choices that people make.

The second topic, human rights, is what connects all of my research together. While few economists explicitly frame their work in terms of human rights, I fundamentally consider myself to be a human rights scholar. A human rights framework leads us to focus on all of the core qualities that make life worth living — or, in other words, everything that contributes to what economists call welfare. More specifically, owing to both the intrinsic and instrumental value of rights, an approach grounded in human rights provides a powerful vehicle for the enhancement of welfare. When rights function intrinsically, the act of exercising or realizing the rights themselves adds to human well-being. When rights function instrumentally, they serve as a means to secure others rights, which, in turn, help to expand well-being. Altogether, welfare emanates from the fulfillment of the human rights that underlie it. In this fashion, welfare is an inherently interdisciplinary concept, arising from a range of domains, from politics to health to education to family to the environment. These domains constitute

the many dimensions of life, all of which are associated with rights.

Thus, the core idea behind welfare — which encompasses the diverse dimensions of life — is a holistic focus on people. Drawing on this perspective, we must always place people at the center of society, as exemplified by what I call “people-centered development” in Chapter 2. Although this point might sound self-evident, we all too often forget about people, even though humanity should be the principal driver of all of our work. In particular, this grounding is essential in the study and pursuit of development, which can be thought of as another term for welfare enhancement. By bringing an explicit focus on people to development, we remind ourselves that development must be about people in both its processes and outcomes. Specifically, the processes of development necessitate both an understanding of human nature and the utilization of human capabilities, while the outcomes of development necessitate the achievement of human welfare. Furthermore, by emphasizing a people-centered approach, a human rights framework provides a universal perspective on development that transcends national boundaries — for human rights hold constant everywhere.

Broadly speaking, the first three chapters of the dissertation pertain to the right to health, while the fourth chapter pertains to the right to education. However, there is much more to the story than that. Chapters 1 and 2, for example, also address political rights: the right to be represented, the right to participate, the right to self-determination, the right to hold leaders accountable.<sup>2</sup> In the context of the research presented therein, political rights carry both intrinsic and instrumental value. Intrinsically, the very act of exercising political rights adds to human freedom and well-being. Instrumentally, political rights serve as a means for securing health rights. In this regard, it is one matter to encode rights, such as the right to health, in the law — but it is another matter to operationalize the law to ensure the realization of such rights. This latter task is the role of political rights; and it constitutes the essential global challenge at the crux of Chapters 1 and 2.

Providing another perspective on human rights, Chapter 3 is situated in a complex social setting in which many factors may impact health outcomes — for instance, the right to knowledge and information and the right to gender equity. Without proper knowledge,

---

<sup>2</sup>From a philosophical standpoint, the research addresses the social contract between citizens and leaders, a notion that dates to Jean-Jacques Rousseau in 1762.

women may be unable to make the best choices in the interest of their health. Without equal rights at home and in society, women may be unable to access the necessary resources to safeguard their health. Lastly, the fourth chapter shows that the right to education is not simply an entitlement to be provided an education; it is also grounded in the ability to *choose* an education. Moreover, as the research shows, the agency to determine one's education level depends on the right to life — because when shortened life expectancy means that the rewards of education cannot be reaped, there is little sense in investing significantly in school. The right to life entails not only the right to enjoy equitable life expectancy regardless of one's identity, but also the right to equal justice under the law, so that inequities in incarceration do not effectively take away one's life.

The third overarching topic of this dissertation is social development. While sometimes, as in the foregoing discussion, I use the more colloquial term of “development,” I like to refer to the subject matter as *social* development for a few reasons. First, “development” alone tends to bring an international connotation. However, development must be a universal concept, because we all remain in a constant process of growth and betterment, not to mention that many wealthy countries contain communities that mirror those in poor countries. Indeed, while the labels carry some utility, I do not subscribe to the categorization of certain countries as “developing” and others as “developed.” Second, at an individual level, the term social development, much like a human rights perspective, helps to make clear that development encompasses outcomes related to all of the social dimensions of life: health, education, civic participation, and the environment, to name a few. Third, development is still about much more than the social dimensions and welfare of an individual alone — for it is also an inherently *social* process, rooted in the connectedness between all people in society. Ultimately, development aggregates the collection of individual achievement, and the synergies created by this achievement, to foster improved life for all of society.

Putting all three major topics of the dissertation together, development can be thought of as both the process and outcome of actualizing human rights, while health constitutes the specific subject matter of my work. But there is yet one more theme that cuts across all of these areas: quality. It is one matter to speak of lofty ideas, but what matters most is how we put these ideas into action. Often, in the world of policy and development, the

difference between high- and low-quality management means the difference between life and death. In this regard, the very process that we use to achieve the ends of development must be a foundational area of focus. Toward this end, Chapter 5 presents key lessons about quality and management, drawing on the years I have spent in the international aid sector and the observations I have made in the course of this dissertation research. At its essence, quality is about integrity — the trustworthiness and faithfulness of the policies, products, and services that permeate our lives. Integrity, however, tends to be a scarce and elusive resource, especially in the present era. Coming across integrity often feels more like a stroke of luck than anything else. How, then, if we want to build a better society, can we *systematize* integrity? How can we weave the incentives for the production of integrity throughout every element of our institutions? It is in this light that Chapter 5 explicates important dynamics in management to help us develop a tangible framework for conceiving of quality in the development sector and elsewhere.

From a geographic standpoint, as much of my dissertation research has been situated in Uganda, it also serves to emphasize a key tenet of my work: that context matters. Fundamentally, global development must be about identifying universal principles and then adapting those principles to every local context. Given all of the time I have spent in Uganda, I have been able to tailor my research to the Ugandan setting in this fashion. Chapter 2, as the culmination of years of thinking about the core economic properties of local political dynamics, is most illustrative of this theme.

Finally, the chapters employ a range of research methods, from a randomized experimental approach in Chapters 1 and 2 to a quasi-experimental observational study in Chapter 3 to a mathematical modeling approach in Chapter 4. Chapter 5 takes a step back from traditional research in applied microeconomics to distill key intellectual observations. Altogether, the diverse methods of the dissertation help to demonstrate the many paths that we can take to knowledge.

## TABLE OF CONTENTS

DEDICATION . . . . .	ii
ACKNOWLEDGEMENTS . . . . .	iii
PREFACE . . . . .	ix
LIST OF TABLES . . . . .	xv
LIST OF FIGURES . . . . .	xvii
LIST OF APPENDICES . . . . .	xviii
ABSTRACT . . . . .	xix
<b>CHAPTER 1 Encouraging Democratic Participation via Community Meetings: Evidence from a Randomized Evaluation of Nudges in Uganda</b>	
1.1 Introduction . . . . .	1
1.2 Setting . . . . .	5
1.3 Empirical Approach . . . . .	12
1.4 Results . . . . .	15
1.5 Discussion . . . . .	16
<b>CHAPTER 2 Governance, Citizenship, and Accountability: A Pilot Study of People-Centered Development in the Ugandan Health Sector</b>	
2.1 Introduction . . . . .	23
2.2 Study Objectives . . . . .	29
2.3 Literature Review . . . . .	29
2.4 Methods . . . . .	31

2.5	Results . . . . .	42
2.6	Preliminary Pre-Analysis Plan for At-Scale Study . . . . .	51
2.7	Discussion . . . . .	55
<b>CHAPTER 3</b>	<b>Demand-Side Factors in Maternal Health Outcomes: Evidence from a Community Health Worker Program in Uganda</b>	
3.1	Introduction . . . . .	62
3.2	Program Background and Data Source . . . . .	67
3.3	Empirical Approach . . . . .	72
3.4	Results . . . . .	77
3.5	Discussion . . . . .	94
<b>CHAPTER 4</b>	<b>Disparity Breeds Disparity: The Impact of Racial Inequalities in Incarceration on Educational Choice</b>	
4.1	Introduction . . . . .	102
4.2	Literature Review . . . . .	107
4.3	Methods . . . . .	108
4.4	Results . . . . .	115
4.5	Discussion . . . . .	122
<b>CHAPTER 5</b>	<b>Quality Production in the Development Sector</b>	
5.1	Introduction . . . . .	127
5.2	Quality as Properly Aligned Incentives . . . . .	129
5.3	Quality Over the Production Cycle . . . . .	135
5.4	Quality through Selection . . . . .	139
5.5	Institutions as a Determinant of Labor Quality . . . . .	141
<b>APPENDICES</b>	. . . . .	146

## LIST OF TABLES

Table 1.1	Analysis of Pre-Intervention Balance Across Treatment and Control Groups . . . . .	13
Table 1.2	Intention-to-Treat Effect Estimation for Any Encouragement Message . . . . .	16
Table 1.3	Intention-to-Treat Effect Estimation for Message Type . . . . .	17
Table 2.1	Meeting Attendance and Participation Summary Statistics . . . . .	47
Table 2.2	Predictors of Meeting Attendance by Households . . . . .	48
Table 3.1	Analysis of Sample Attrition . . . . .	75
Table 3.2	Pre-Program Falsification Tests . . . . .	76
Table 3.3	Intention-to-Treat Effect Estimation . . . . .	78
Table 3.4	Pre-Program Falsification Tests for Home vs. Neighboring Village CHW Treatment . . . . .	81
Table 3.5	Pre-Program Falsification Tests for Male vs. Female CHW Treatment	82
Table 3.6	Intention-to-Treat Effect Heterogeneity Based on CHW Location . . .	83
Table 3.7	Intention-to-Treat Effect Heterogeneity Based on CHW Gender . . . .	84
Table 3.8	Intention-to-Treat Effect Heterogeneity Based on Hours to Facility . .	86
Table 3.9	Intention-to-Treat Effect Heterogeneity Based on Gestational Age . . .	88
Table 3.10	Program Effects for Primigravida Women. . . . .	90
Table 3.11	Recipient Assessment of CHW Visits . . . . .	92
Table 3.12	Pre- to Post-Intervention Changes for Entire Sample . . . . .	93
Table 3.13	Predictors of Post-Program Outcomes . . . . .	95
Table 4.1	Remaining Lifetime at Age 18 for Black and White Men, with and without Incarceration Adjustment . . . . .	115
Table 4.2	Predicted Change in Years of Schooling Across Net Interest Rate and Assumed White-Male Elasticity . . . . .	116
Table 4.3	Predicted Change in Years of Schooling Across Net Interest Rate, Assumed White-Male Elasticity, and Education-on-Incarceration Gradient Weight . . . . .	121
Table 4.4	Predicted Change in Years of Schooling Across Net Interest Rate and Schooling Lower Bound . . . . .	122
Table 4.5	Predicted Change in Years of Schooling Across Elasticity Methods and GED Coding . . . . .	123
Table C.1	Pooled Pre-Program Falsification Tests . . . . .	179



Table C.2	Pre-Program Time Trends Across Treatment and Control Groups . . .	180
Table C.3	Pre-Program Time Trends Based on CHW Location . . . . .	181
Table C.4	Pre-Program Time Trends Based on CHW Gender . . . . .	182
Table C.5	Analysis of Pre-Intervention Balance Across Treatment and Control Groups . . . . .	184
Table D.1	Imputed Education Years for ACS Data from 2008 and After . . . . .	186
Table D.2	Imputed Education Years for ACS Data from Before 2008 . . . . .	186

## LIST OF FIGURES

Figure 1.1	Sampling and Randomization Flow Chart . . . . .	9
Figure 2.1	Experiment Randomization Matrix . . . . .	31
Figure 4.1	Male Education Trends by Race . . . . .	103
Figure 4.2	Male Survival and Incarceration by Race . . . . .	105
Figure 4.3	Estimated Schooling Differences for Various Elasticities . . . . .	117
Figure C.1	Pre-Treatment Four ANC Visits Trends . . . . .	177
Figure C.2	Pre-Treatment Skilled Delivery Trends . . . . .	177
Figure C.3	Pre-Treatment Breastfeed 30 Minutes Trends . . . . .	178
Figure C.4	Pre-Treatment Improper Cord Care Trends . . . . .	178
Figure D.1	Institutionalization Rates by Type for Men in 1980 . . . . .	188

## LIST OF APPENDICES

APPENDIX A: Appendix for Chapter 1 . . . . .	146
APPENDIX B: Appendix for Chapter 2 . . . . .	165
APPENDIX C: Appendix for Chapter 3 . . . . .	175
APPENDIX D: Appendix for Chapter 4 . . . . .	185

## ABSTRACT

Spanning diverse perspectives, the five chapters of this dissertation relate to health, human rights, and social development. Broadly speaking, the first two chapters center on the supply side of healthcare provision in Uganda, while the third chapter, also situated in Uganda, focuses on demand-side barriers to health. Complementing this work on healthcare delivery and health behavior, Chapter 4 builds on the role of health in human capital formation. Finally, Chapter 5 zooms out from the research questions of Chapters 1-4 to present key observations about the development sector.

Specifically, Chapter 1 focuses on citizen participation, which is considered a key priority in making public service delivery more accountable. I conduct a randomized controlled trial to evaluate the effect of differently themed messages encouraging households to attend community meetings that address local healthcare quality. I find that receiving a message — of any type — significantly increases meeting participation, but that the specific type of message does not appear to induce any further effects. The results provide important evidence on how to better foster democratic engagement.

Building on this work, Chapter 2 addresses the widespread accountability gaps in the public health sectors of low- and middle-income countries. Focusing on the role of political leaders in promoting improved health sector performance, I present the results of a pilot study of the following interventions: (a) citizen reporting meetings with local politicians on health service delivery and (b) politician skills training on monitoring local government health centers. The study is one of a few to evaluate programs directly targeting political economy inefficiencies. Both interventions achieve substantial stakeholder engagement, laying the groundwork for an at-scale evaluation.

Using a difference-in-differences design, Chapter 3 evaluates a community health worker program aimed at improving maternal health. I find relatively little evidence of an overall

program effect on health behaviors, although heterogeneity analysis provides suggestive evidence of treatment effects. In contrast to these weak findings, I find large improvements in healthcare-seeking behavior between the pre- and post-intervention periods. Altogether, the study helps to fill a key evidence gap on a common intervention.

Chapter 4 turns to an under-recognized pathway through which mass incarceration may affect social welfare: by burdening young Black men with substantial *expectations of future incarceration* and thereby altering their life decisions before they even experience any imprisonment. Together with Hoyt Bleakley, I mathematically model and empirically quantify the effects of excess incarceration and mortality on educational choice across Black and White men in the U.S. We find that this future expectations pathway explains most, if not all, of the persistent racial education gap in men. This finding has significant implications for policies aimed at addressing racial disparities in education.

Lastly, drawing on the dissertation's running theme of quality, Chapter 5 builds an intellectual framework for quality production in the development sector. In the context of development, quality is an especially complicated object for two reasons: (1) the very attempt to measure quality often undermines quality; (2) efforts to properly define quality at a product level may be infeasible. Based on these observations, I demonstrate the importance of a more upstream perspective in assessing quality in the development industry. In doing so, I establish a mathematical model of production cycle quality, while also elucidating the roles of selection and institutions in promoting quality in the aid sector.

# CHAPTER 1

## Encouraging Democratic Participation via Community Meetings: Evidence from a Randomized Evaluation of Nudges in Uganda

### 1.1 Introduction

Beyond the mere provision of needed resources to support public services, the global development sector has increasingly emphasized the imperative of building effective institutions to encourage the responsible use of existing resources. Toward this end, many scholars and practitioners believe that citizen participation in local affairs enhances the accountability of public service delivery, which is recognized as a key priority in the fields of governance and development ([World Bank 2003; 2017](#)). The rationale for citizen participation often centers on the observation that citizens, as users of public services, ought to have superior information about service quality as well as more favorably aligned incentives for improving those services. These notions parallel the tenets underlying democracy, which hold that citizens — based on their best interests — should have the right to determine their political leaders.

To ultimately be successful in impacting public service delivery, grassroots programs must, as a prerequisite, produce substantial ‘first-stage’ participation effects. Along these lines, a common priority is to ensure that participation is representative of the target population, so that public sector performance will be responsive and accountable to the population’s unique composition. In particular, in settings that are often characterized by elite capture, fostering participation by a broad section of the populace can improve social welfare by helping to create more equitable public policy.<sup>1</sup> With this goal in mind, this study aims to deepen

---

<sup>1</sup>To the contrary, it is also theoretically possible that enhanced participation could lead to counterproductive effects due to factors such as private incentives, incomplete information, and disruptive public behavior. Ultimately, the consequences of these possibilities likely depend on the nature and management of the participatory program in question. In the present study, which involves public meetings (as described further below), the meeting facilitation is carried out according to professional, standardized guidelines in

our understanding of strategies to encourage strong citizen participation, an important input in governance across a variety of contexts.

While a number of studies have evaluated the effects of participatory programs on public service provision outcomes (Casey 2018; Banerjee et al. 2010; Björkman and Svensson 2009; Olken 2007), little research in economics examines strategies to increase participation itself, which, like the public services that programs seek to improve, is generally a public good. A lone piece of evidence on this topic comes from Olken (2007), who, in the context of an accountability meetings intervention in Indonesia, finds that the distribution of meeting invitations to households induces significant grassroots participation. Beyond the development literature, efforts to induce community participation also relate to the literature on choice architecture — and on how behavioral nudges can encourage more optimal decision-making.<sup>2</sup> In particular, in settings with inefficiently low contributions toward public goods,<sup>3</sup> nudges can play an important role in fostering greater prosocial behavior. In this light, nudges have been studied with regard to goals such as promoting charitable and blood donations (Bartke et al. 2017; Zarghamee et al. 2017; Goswami and Urminsky 2016; Stutzer, Goette, and Zehnder 2011), encouraging energy conservation (Brandon et al. 2019; Costa and Kahn 2013), and increasing voter turnout (Dale and Strauss 2009).

Although nudges have demonstrated positive effects on prosocial behavior in many contexts, evidence has also emerged that nudges, especially those that appeal to private incentives, may lead to the unintended consequence of crowding out prosocial behavior (Mellström and Johannesson 2008; Frey and Oberholzer-Gee 1997; Ariely, Bracha, and Meier 2009). Moreover, key questions about the mechanisms and optimal properties of nudges remain. In particular, how important is the specific content of nudges for motivating behavior change? What is the role of individual vs. general targeting in the design of nudges? Do multiple nudges exhibit crowd-out effects?

---

an effort to reduce potentially ineffective or harmful participation. In addition, by seeking to foster consensus across a large, representative group, the meetings should minimize any effects of individual participant behavior related to private incentives or incomplete information. Please see Chapter 2 for more detail on the implementation of the meetings themselves. Finally, holding all else equal, citizens with stronger private incentives in particular may have a higher propensity to participate in local affairs; therefore, in the same spirit as countering elite capture, efforts to induce greater participation from the general population should attract the involvement of more citizens with weaker private incentives.

<sup>2</sup>See Thaler and Sunstein (2009) for a broad overview of the literature.

<sup>3</sup>See Fehr and Schmidt (1999) for a compilation of evidence on free riding.

Building on the foregoing observations and questions, this study evaluates a messaging nudge intervention in Uganda to increase citizen participation within the context of a democratic accountability program. Specifically, I use a randomized controlled trial to measure the effect of differently themed messages encouraging households to attend community accountability meetings between citizens and political leaders that address the quality of local health service provision. Since democratic participation is a public good, attendance at such citizen-politician meetings is subject to free-rider effects. At the same time, since individual citizens (particularly non-elite citizens) in the political system are akin to price-takers in a competitive goods market, the complaints of one citizen alone are unlikely to influence the behavior of a political leader. Taken together, these features make collective action an essential element of political processes. Therefore, this study seeks to shed light on how we can more effectively foster participation for the purpose of collective political action. To the extent that such collective action leads to improvements in social welfare, low-cost nudges encouraging civic engagement may have high returns.

Importantly, citizens may face several constraints to political participation, falling into four main categories: (1) organization, (2) information, (3) transportation, and (4) motivation. First, citizens likely require robust leadership to help provide structure for collective organization and to offer them a concrete course of action. Without such leadership, citizens may be unlikely to participate. In a perfectly functioning democracy — with perfect information flow between citizens and politicians and complete accountability of political leaders — we might expect the politicians themselves to provide the necessary leadership and resources to organize meetings with their constituents. In imperfect democracies or other political systems, third-party intervention may be required to invigorate collective action. Second, in addition to these organizational constraints, information deficiencies may limit citizen participation. Informational gaps may be as basic as a lack of awareness regarding opportunities to participate in collective action, which citizens may need for the purposes of advance planning. Citizens may also experience incomplete information regarding their political rights and entitlements. Third, difficulty accessing transportation may create a physical barrier to citizen participation. Especially in large jurisdictions where major political events are typically held in the center, citizens from more distant areas may be inhibited



from participating. Lastly, participation may be limited by motivation. Even when citizens have all of the necessary information and resources to adopt an accessible course of action, they may need additional persuasion to inspire them to join a collective action effort.

In this study, the community accountability meetings are organized by a non-governmental organization, thereby eliminating leadership and organizational constraints. In addition, the implementation of the meetings is accompanied by substantial community-wide mobilization efforts, which serve to alleviate informational barriers to participation. The meetings also take place in two locations within each governing jurisdiction, helping to make the events as accessible as possible. Therefore, by testing different encouragement messages to increase attendance at the meetings, this study focuses on motivational constraints to participation. Specifically, I examine the following three themes for the messages: (1) public duty, (2) private benefits, and (3) civic participation. Along with evaluating the impact of each specific message theme in comparison to a neutral message, I also measure the effect of receiving any message at all.

I find that receiving a message — of any type — nearly doubles household participation in the meetings, but that the specific type of message does not appear to induce any further effects. In other words, an encouragement message produces a substantial first-order effect, but the marginal characteristics of a message correspond, if anything, to marginal effects. The findings imply that the messages act as substitutes, and that none of the specific message themes, including the private benefits theme, crowd out prosocial behavior. Moreover, the findings — along with the fact that the concurrent community-wide mobilization activities likely created widespread awareness of the meetings throughout the study catchment area — suggest that the mechanism underlying the observed effect is the individualized nature of the encouragement messages. The results come with an important caveat, however: all households that were assigned a message were also assigned to be surveyed, while all households that were not assigned a message were not assigned to be surveyed. Therefore, the effect of message assignment cannot be formally distinguished from the effect of survey assignment. Nevertheless, regardless of the precise source of the observed effect, the effect remains causal and helps to contextualize the null effect associated with the particular message theme.

The remainder of this article is structured as follows. In Section 1.2, I describe the

background of the community accountability meetings program, its relevant features to the encouragement message study, and the data source. Then, in Section 1.3, I outline the study's empirical approach. Section 1.4 presents the results, and Section 1.5 discusses the significance of the findings.

## 1.2 Setting

### 1.2.1 Program Background

The evaluation of the encouragement messages is situated within the Community-Centered Development Program, a pilot governance initiative carried out by Progressive Health Partnership (PHP), a non-profit organization, in collaboration with the Office of the Prime Minister (OPM) in Uganda. Under the program, PHP is carrying out a series of quarterly feedback, reporting, and accountability meetings between citizens and local politicians to discuss the quality of local public healthcare delivery. Fundamentally, the meetings aim to give citizens a platform from which they can express complaints to their leaders, exercise their rights, and work collectively to better influence government service delivery.

Specifically, PHP carries out the meetings in ten randomly selected subcounties/town councils/divisions in the Ankole Region of Uganda. As equivalent administrative units across rural, periurban, and urban areas, respectively, each subcounty, town council, and division has a government health center referred to as a Health Center III and is headed by a Local Council III (LC3) Chairperson, a democratically elected leader who presides over the council that governs the jurisdiction. Broadly speaking, the legal authorities of LC3 chairpersons include overseeing the performance of government employees and the provision of government services within their jurisdictions ([Republic of Uganda 2000](#)). In the health sector, chairpersons have several tools at their disposal to achieve these ends, such as more closely monitoring health facilities, strengthening the local Health Unit Management Committee, and adopting professional or social sanctions against health workers.

Creating opportunities for regular feedback and reporting, PHP's meetings bring together citizens with the LC3 Chairperson to discuss the quality of service delivery at the local Health

Center III. Based on citizen complaints, the LC3 Chairperson can then adopt monitoring or enforcement measures to help improve services at the health center. In each administrative unit, PHP carries out two meetings per quarter — one in the parish/ward that contains the Health Center III and another in a randomly chosen parish/ward to help increase intervention coverage and to make the meetings as accessible as possible.

Prior to each meeting, PHP implements a multipronged, community-wide mobilization process aimed at fostering widespread, representative participation. At the inception of the program, PHP visited all of the involved communities to inform local stakeholders of the planned activities, solicit their assistance in the meeting mobilization process, and recruit a community health worker who is expressly tasked with serving as a local mobilizer. The specific mobilization procedures for each meeting then consist of the following. First, 14 days prior to the meeting, the PHP implementation team calls the LC3 Chairperson, all of the village chairpersons within the meeting parish/ward, and the local mobilizer to remind them of the upcoming activities and to request that they begin informing community members of the meeting. Seven days prior to the meeting, the local mobilizer delivers general invitation letters to various religious and opinion leaders, requesting their assistance in announcing the upcoming meeting during community gatherings. The local mobilizer also distributes and hangs announcement posters in key areas throughout the community. PHP again calls all local leaders to remind them of the upcoming activities and also sends text message reminders to them. Three days prior to the meeting, PHP carries out in-field mobilization by driving throughout the target community with a mobile public address system to make announcements about the upcoming meeting. Two days prior to the meeting, PHP executes additional reminder calls and text messages for all local leaders. Finally, in the morning of the day of each meeting, PHP carries out additional in-field mobilization using the public address system. Copies of all of the key materials used in the community-wide mobilization activities appear in [Appendix A.3](#).

Altogether, PHP’s activities mean that all households in the target communities are exposed to significant mobilization efforts through the meetings program. Layered on top of these community-wide mobilization activities, the encouragement messages focus on alleviating motivation-related constraints to participation, as discussed in [Section 1.1](#). Specifically,

for the first round of meetings that PHP carried out, between January and March 2020, I randomized households across four encouragement message arms: (1) neutral, (2) public duty, (3) private benefits, and (4) civic participation. Households in the neutral group receive a basic message encouraging them to attend the meeting in their community. The messages for the three remaining groups, which incorporate content relevant to each respective theme, all build very precisely on the text of the neutral message to maintain maximal uniformity and to eliminate the potential generation of other confounders in the alterations to the message text. In brief, the public duty message emphasizes the responsibility of citizens to participate in the meeting for the sake of overall community welfare. The private benefits message stresses that the household will benefit from the improvements in health services that the meetings achieve. The civic participation message focuses on the meeting as an opportunity for different social groups and under-heard citizens to raise their voices with the LC3 Chairperson.

In practice, each message consists of several components, rather than simply a single message. First, between October and November 2019, at the time of the baseline survey for a separate evaluation of the pilot meetings program that I conducted, the interviewer administering the questionnaire delivered an extended oral message according to the randomly assigned arm of each household. Prior to this element of the intervention, interviewers were trained to strictly adhere to the text of each message, to read all messages with consistent enthusiasm, to avoid altering the tone of their delivery based on the message theme, and to answer any questions from the respondent in an objective fashion that did not stray from the core content of the message. Following the verbal exchange, the interviewer left behind a written copy of the message with the household. Along with the more textual version of the message provided to the respondent, the interviewer also distributed a themed poster that emphasized the message in a more visually appealing format. Specifically, the posters depicted three individuals discussing the importance of the upcoming meeting, with the text of their conversation tailored to the particular message theme. Finally, households that provided contact telephone numbers during the baseline survey received themed text messages seven days and two days prior to the meeting in their community. Copies of all of the materials used for the encouragement messages appear in [Appendix A.2](#).

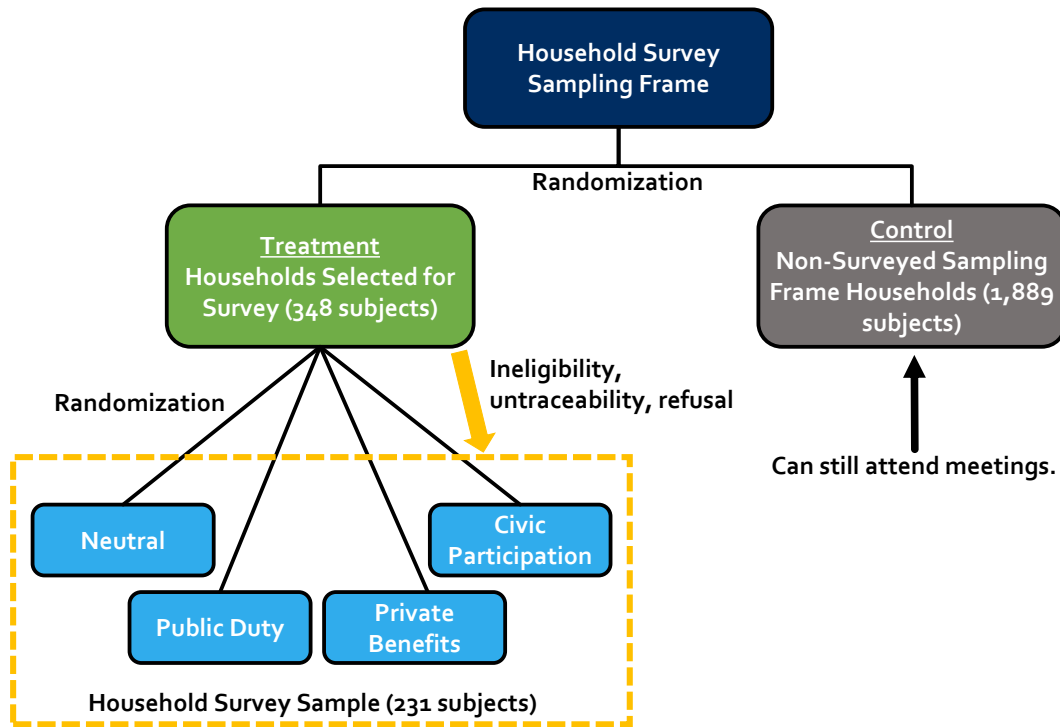
## 1.2.2 Data Collection

As mentioned above, for a separate evaluation of the pilot meetings program, I carried out a baseline survey of 231 households in the intervention catchment area. I also carried out a meetings survey that recorded observations about the conduct of each community accountability meeting. The data collection for the encouragement message evaluation occurs within the context of these two surveys, as meeting attendance by households is the outcome of interest. Overall, the data collection is geared toward collecting information to measure the effect of receiving (1) any message at all and (2) a specifically themed message.

Figure 1.1 provides a diagram of the sampling and randomization process. In advance of the baseline household survey, PHP’s data collection team carried out enumeration of all households in the area to form the sampling frame. To complete the household enumeration process, the data collection team visited the village chairperson in each of the sample villages to request a copy of the village’s household registry. Members of the data collection team then reviewed the registry with the chairperson, verifying that the list of households was comprehensive, accurate, and up-to-date. To be eligible for the survey, households were required to have one or more children under five as well as a respondent above 18 years of age. Therefore, as an initial eligibility screen, the data collection team asked the chairperson whether each household contained one or more children under five. In general, village chairpersons are knowledgeable about the households in their community. However, the question responses included a “Don’t know” option. All households with “Yes” or “Don’t know” responses for the child under five question were included in the sampling frame. In all cases, eligibility was confirmed directly with the household at the time of survey administration.

From the sampling frame, I then randomly selected 348 households for the baseline survey of the pilot meetings program (the green box in Figure 1.1). Following the selection of the survey sample, I further randomly assigned the chosen households across the four encouragement message arms. Therefore, all households selected for the survey sample were assigned to an encouragement message. At the same time, another group of 1,889 households remained in the sampling frame: those that were not randomly chosen for the survey and therefore did not receive any form of encouragement message, but that could still decide to

FIGURE 1.1: SAMPLING AND RANDOMIZATION FLOW CHART



*Notes:* This figure diagrams the sampling and randomization process. For the evaluation of the main message effect — that is, the effect of being assigned any message — the treatment group represented by the green box is compared to the control group represented by the grey box. For the evaluation of the message type, each themed treatment arm amongst the blue boxes is compared to the neutral message group.

attend one of the community accountability meetings (the grey box in Figure 1.1). I use this group of households as a control arm to evaluate the effect of receiving any message at all, while I use the neutral message group as a comparison to evaluate the effect of each message theme.

The collection of meeting attendance outcomes involved two different elements of data collection. First, as part of the meetings survey, I gathered attendance records for each meeting, recording the name, associated household, address, and phone number of each meeting attendee. PHP’s data entry team then undertook the process of matching (a) the attendance records for meeting attendees reported to be from households in the catchment area with (b) the list of both eligible and ineligible households created during the baseline household survey sampling frame enumeration process. The matching process used this full list of eligible and ineligible households — generated before implementing the initial eligibility screen to trim the list to the final sampling frame — since any household in the

community, regardless of eligibility status, was free to attend the meetings, and since the meeting attendance records contained no eligibility information. Overall, using this full list helped to make the matching process as seamless as possible.<sup>4</sup> Later, at the analysis stage, I restrict the estimations to the sampling frame households only, as depicted in the green and grey boxes of Figure 1.1. Thus, altogether, by merging datasets (a) and (b) together, I am able to link the meeting attendance outcomes with households in the sampling frame.

The matching process was an intensive effort involving double entry along with multiple rounds of review and reconciliation to ensure maximal fidelity of the data merge. For records that were difficult to match, the data entry team went to significant lengths to obtain more information by attempting to contact attendees via their provided phone numbers or, if a phone number was not available, through other contacts in the community. Finally, the matching procedure was conducted in an entirely uniform fashion for all households, regardless of their membership in the household survey sample (and therefore receipt of an encouragement message). While carrying out the matching procedure, the data entrants were blinded to the survey status of each household, thereby ensuring consistency in the quality of the matching process across all households. Ultimately, of a total of 1,051 meeting attendees who came from households in the catchment area, 896 could be matched to the sampling frame. For the remaining attendees who could not be matched, there may be several explanations. In some instances, due to inaccurate knowledge of official village boundaries, attendees may have misreported their home village as one of the units in the catchment area; they would therefore be included in the matching process but could not be located in the sampling frame. In other cases, attendees may have reported a different household head, or an alternative name for the household head, compared to the information listed in the sampling frame. Despite the substantial inquiry made in an effort to resolve such discrepancies, these cases were sometimes impossible to match. Overall, as the causes of the failed matches are relatively idiosyncratic, and as there is no reason to believe that

---

<sup>4</sup>Using the full list also provided more accurate information on the match failure rate, which I discuss further below. If the matching process had only used the list of eligible sampling frame households, we would have failed to match many meeting attendees simply because they came from ineligible households and did not appear in the reference list, rather than because of a more substantive data quality issue. In addition, using the full list produced the added benefit of generating data for ineligible households as well, although those data, apart from contributing to the initial match results, are not used in this article.

the failed matches disparately affected one study group or the other, the unmatched meeting attendees do not create major generalizability concerns.

Second, in addition to the matched records, I collected meeting attendance data specific to the households in the baseline survey sample that were assigned the different message themes. At the time of the survey administration, the interviewer gave the respondent a sheet of paper containing the household's ID and requested that the respondent or another household member bring the paper to the upcoming community meeting, should the household choose to attend. Then, at the time of the meeting, the program implementation team asked whether any community members in attendance had been visited for an interview by the data collection team. Those who brought their ID sheets to the meeting then submitted the sheets. In case a household reported having been interviewed, and therefore receiving an encouragement message, but lost the ID sheet, the program implementation team collected thorough information on the household's identity so that the household's presence in the survey sample could later be verified. The data entry team then recorded the collected household ID's and verified those households that reported having been visited for an interview but did not have an ID sheet. Of the 96 households from the survey sample that were confirmed to have attended a community meeting, 69 households submitted an ID sheet, while the remaining 27 had to be identified manually after reporting that they had lost their ID sheets. While these 27 households were confidently verified, an additional six households that attended the meetings reported having received an interview and encouragement message but could not be confidently matched to the survey sample. This group of households may have misrepresented their inclusion in the survey. On the whole, I consider the data specific to the survey sample to be higher-quality than the data from the sampling frame matching process. However, when completing the matching process for all households, the data entry team did not have access to any of the sample-specific meeting attendance data.



## 1.3 Empirical Approach

### 1.3.1 Analysis of Main Message

To measure the effect of being assigned any encouragement message, I use a simple difference-in-means test between households that were and were not assigned a message (the green versus the grey boxes in Figure 1.1). The resulting estimate of the message effect is an intention-to-treat (ITT) effect because households that were assigned to receive a message did not necessarily receive one. For example, some households were chosen for the survey from the sampling frame but then could not be located, were deemed ineligible, or refused to participate at the point of data collection. In such cases, the household would not receive an encouragement message.

Since the control group used for the measurement of the main effect of the encouragement messages is composed of the non-surveyed households from the sampling frame, no covariate data are available for control households, and therefore it is not possible to verify balance between the groups. However, because households were randomly selected to receive a message, the treatment is orthogonal to the potential outcomes, and the identification strategy is internally valid for detecting a causal treatment effect. Nevertheless, because the control households that were not assigned a message also were not included in the survey sample, it is not possible to formally differentiate whether an observed treatment effect is due to message assignment or survey sample inclusion.

### 1.3.2 Analysis of Message Type

To measure the effect of the specific message theme, I compare each of the themed message arms to the neutral message arm. This comparison is given by the following regression specification:

$$Y_i = \beta_0 + \beta_1 D_i + \beta_2 B_i + \beta_3 P_i + u_i, \quad (1.1)$$

where  $Y_i$  denotes the outcome variable of interest for household  $i$ ;  $D_i$ ,  $B_i$ , and  $P_i$  are indicator variables for the public duty, private benefits, and civic participation messages, respectively;

and  $u_i$  is an error term.

TABLE 1.1: ANALYSIS OF PRE-INTERVENTION BALANCE ACROSS TREATMENT AND CONTROL GROUPS

	Neutral	Public Duty	Private Benefits	Civic Participation
Satisfaction with Health	3.507	3.565 (0.814)	3.722 (0.377)	3.417 (0.748)
Satisfaction with Community	4.045	3.919 (0.588)	4.167 (0.594)	3.854 (0.465)
Can Influence Local Gov.	0.697	0.710 (0.876)	0.685 (0.891)	0.833* (0.086)
Can Improve Gov. Health Services	0.723	0.806 (0.271)	0.741 (0.830)	0.875** (0.042)
Attended Village Meetings	0.477	0.516 (0.662)	0.509 (0.728)	0.553 (0.430)
Attended LCIII Meetings	0.134	0.129 (0.930)	0.167 (0.626)	0.083 (0.383)
Community Join Together	0.703	0.738 (0.670)	0.765 (0.460)	0.750 (0.584)
Community Pressure Chairperson	2.970	3.081 (0.582)	2.925 (0.827)	2.851 (0.593)
Information on Healthcare	0.500	0.597 (0.275)	0.704** (0.023)	0.542 (0.663)
Drugs and Supplies Available	0.697	0.661 (0.671)	0.660 (0.675)	0.587 (0.239)
Staff Normally Attend Work	0.894	0.864 (0.617)	0.904 (0.860)	0.957 (0.203)
Time to Health Center	51.106	51.933 (0.938)	53.593 (0.814)	46.196 (0.590)
Num. Health Center Visits	3.879	3.710 (0.811)	3.212 (0.307)	3.667 (0.794)
Num. Visits	2.288	2.806	2.296	2.667

Other Providers		(0.437)	(0.987)	(0.551)
Num. Household Members	5.701	6.242 (0.211)	5.759 (0.888)	5.688 (0.977)
Respondent Highest Grade in School	6.806	6.226 (0.336)	6.407 (0.554)	6.146 (0.313)
HH Head Highest Grade in School	5.136	6.056 (0.142)	5.962 (0.219)	6.907** (0.012)
Economic Index	0.067	-0.022 (0.786)	-0.216 (0.423)	0.179 (0.760)
Missing Any Covariate	0.194	0.226 (0.661)	0.167 (0.699)	0.208 (0.852)
Joint test ( $\chi^2$ ): PD	0.628			
Joint test ( $\chi^2$ ): PB	0.669			
Joint test ( $\chi^2$ ): CP	0.121			

*Notes:* This table displays the results of balance tests between each themed message arm and the neutral message arm. The mean of each variable is shown for each study arm. Underneath each themed message group mean, the  $p$ -value from a difference-in-means test between the themed group and the neutral group is shown in parentheses. The bottom panel of the table shows  $p$ -values from joint tests of differences for each themed group relative to the neutral group. All variables are measured at the household level, and all tests use heteroskedasticity-robust standard errors. \*Significant at 10% level, \*\*Significant at 5% level, \*\*\*Significant at 1% level.

Since this aspect of the analysis is confined to surveyed households, I am able to use the sample-specific attendance data for the meeting attendance outcome variable. In addition, I test for balance between each themed message arm and the neutral message arm, according to a variety of baseline dimensions. Table 1.1 displays the results, showing  $p$ -values from difference-in-means tests in parentheses below each themed message group variable mean. Only a few baseline characteristics are statistically significantly different between the themed groups and the neutral group. For the private benefits message group, whether the respondent reports having sufficient information to be able to evaluate the quality of services at the local government health center is significant at the 5-percent level. For the civic participation group, two characteristics related to attitudes toward capabilities in community affairs exhibit differences significant at the 10-percent and 5-percent level, respectively: whether the respondent feels that people like him/her (1) are able to influence the decisions made by the local government and (2) have the power to improve how the government provides healthcare to the local community. In addition, the household head education level is higher for this

group compared to the neutral message group ( $p < 0.05$ ). The bottom of the table displays the  $p$ -values from joint tests of differences for each group. Despite some of the differences in individual variables, none of the joint tests demonstrate a statistically significant difference. Altogether, the balance shown between the themed message groups and neutral message group supports the internal validity of the randomized experiment identification strategy.

## 1.4 Results

Table 1.2 shows the results for the estimation of the effect of being assigned to any encouragement message. Households in the treatment group experience a 15.3 percentage-point increase in the likelihood of attending a meeting, an effect that is highly statistically significant. Thus, given the average meeting attendance level of 16.6 percent in the control group, treatment nearly doubles the probability of household attendance at meetings. In addition to this extensive margin effect — on the binary of sending a household representative to a meeting — the treatment induces an intensive margin effect. Conditional on attending a meeting, households assigned to treatment also send more household members to the meetings. This result is demonstrated by the tests displayed in the second and third lines of the table, which I perform only among households that attended a meeting. Treatment households are 6.8 percentage points, or 51.9 percent, more likely to send more than one household member to meetings, although this effect is barely insignificant at the 10-percent level. If I instead use the number of household members sent to the meeting as the outcome variable, the effect is statistically significant at the 10-percent level, but the effect size of 0.086 is substantively insignificant. In total, of the 425 households across the treatment and control groups that were represented at the meetings, 14.8 percent sent more than one household member to a meeting, with the majority of those households sending two members.<sup>5</sup>

While being assigned to receive any message has a significant effect on meeting attendance by households, the specific type of message appears to make no difference. Table 1.3 displays the results from the estimation of the effects of each themed encouragement message, relative

---

<sup>5</sup>Recall that the matching of the meeting attendance records with households was done for both eligible and ineligible households. Therefore, the figure of 896 matches provided in Section 1.2.2 substantially exceeds the 425 households cited here, which only come from the group of eligible sampling frame households.

TABLE 1.2: INTENTION-TO-TREAT EFFECT ESTIMATION FOR ANY ENCOURAGEMENT MESSAGE

	(1)	(2)	(3)	(4)
	Control Mean	Treatment Mean	Message Effect	$p$ -value
Attended Meeting	0.166	0.319	0.153	0.000
> 1 Household Member	0.131	0.198	0.068	0.112
Number HH Members	1.131	1.216	0.086	0.070

*Notes:* This table displays results for the evaluation of the effect of being assigned to any encouragement message. The attended meeting outcome is measured for all households, while the second and third outcome variables are conditional on a household sending at least one member to a meeting. All variables are measured at the household level, and all tests use heteroskedasticity-robust standard errors.

to the neutral message arm. Although the estimates appear to be imprecise due to limited power, none of the estimates are statistically significant. The private benefits message is associated with a 4.8 percentage-point reduction in meeting attendance, while the civic participation message is associated with a 6.1 percentage-point increase. However, as shown in the second column of the table, the addition of demographic controls to the regression attenuates all of the effect estimates toward zero.

## 1.5 Discussion

This article uses a randomized evaluation to measure the effect of encouragement messages aimed at fostering citizen participation in local governance in Uganda. In contrast to many studies that evaluate the outcomes of participatory programs in relation to services provided by public bureaucrats, this study focuses on the determinants of participation itself in the context of collective political action. Specifically, by delivering encouragement messages to households, the study targets motivational constraints to citizen participation.

While the particular type of message — between themes of public duty, private benefits, and civic participation — does not appear to have an effect on households, I find that receiving any message, regardless of type, nearly doubles the rate of household participation in community accountability meetings. Hence, messages have a significant first-order effect, but their marginal characteristics make little difference. The large size of the first-order effect is

TABLE 1.3: INTENTION-TO-TREAT EFFECT ESTIMATION FOR MESSAGE TYPE

	(1)	(2)
	Attended Meeting	Attended Meeting
Public Duty	-0.015 (0.087)	-0.004 (0.096)
Private Benefits	-0.048 (0.090)	-0.010 (0.098)
Civic Participation	0.061 (0.095)	0.032 (0.103)
Demographic Controls	No	Yes
Mean Control	0.418	0.418
Observations	231	231

*Notes:* This table displays results for the evaluation of the effect of each message theme. Heteroskedasticity-robust standard errors are shown in parentheses below the regression coefficient estimates. The second column includes demographic controls, which include all of the same covariates used in the balance tests shown in Table 1.1, with the exception of the missing any covariate indicator. Please see Appendix A.1.1 for further detail on how missing values are treated. All variables are measured at the household level. \*Significant at 10% level, \*\*Significant at 5% level, \*\*\*Significant at 1% level.

remarkable; to the extent that behavior change by politicians and other governing officials depends on the number of citizens involved in collective political action, the encouragement messages offer an intervention to significantly enhance the ability of citizens to exert pressure on their leaders for improved governance. Moreover, the results of the study have two key theoretical implications. First, the message nudges in this study function as substitutes, as the different messages appear to be equally effective, with their specific content inconsequential. This finding parallels the results of similar studies (Altmann and Traxler 2014; Dale and Strauss 2009). Second, none of the messages, including the message themed on private benefits, exhibit any backfire effects by crowding out prosocial behavior. Altogether, these findings lead to a reassuring conclusion: we can achieve broad appeal with simple messages, without having to worry much about the unique content of the messages.

What underlies the observed effect of the encouragements? Although the mechanism of effect cannot be formally determined, the study findings offer important mechanistic insights about nudges. While the encouragement messages were designed to test motivational con-

straints to participation, the null results for message type suggest that the overall message effect did not operate through a theme-based motivation or persuasion pathway. An alternative explanation is that the messages acted as reminders by alleviating limited attention constraints. However, the message intervention in this study was layered over a set of highly organized community-wide mobilization activities for the meetings. This setup implies a relatively high level of awareness of the meetings throughout the communities, as exemplified by the roughly one in six households in the control group that attended the meetings. Therefore, the limited attention constraint hypothesis is unlikely. Instead, the simultaneous presence of significant background mobilization activities suggests a more likely mechanism: the delivery of the encouragement messages through individual rather than general outreach persuaded many households to attend the meetings. Given that the content of the general mobilization messages (shown in Appendix A.3) was similar to that of the encouragement messages, the individual outreach of the encouragement messages appears to account for the major effect on household behavior. Thus, the encouragement messages may still have operated through a motivational pathway, only that the motivational effect hinged on the individualized nature of the messages rather than their specific content.

The presence of significant community-wide mobilization activities may also have implications for crowd-out effects between nudges. Although it is not possible to draw formal conclusions on crowd-out in this study, we might expect participation-inducing interventions to exhibit diminishing returns. If so, the general mobilization activities might crowd out the effects of the encouragement messages, implying that the estimated effect of the messages may be a lower bound. It is notable that the encouragement messages nonetheless produce such a large effect. In other settings, where general mobilization activities may not be executed with comparable intensity, encouragement messages may induce participation to an even greater extent. Alternatively, it is also possible that no crowd-out effects exist,<sup>6</sup> and that general outreach and individual outreach even act as complements. Overall, the crowd-out question is a key area for future research.

One important caveat to the foregoing findings, however, is that the effect of being as-

---

<sup>6</sup>See [Brandon et al. \(2019\)](#) for an example of energy conservation nudges that do not crowd each other out.

signed to the message treatment cannot technically be distinguished from the effect of being assigned to the survey. Since the control group for the analysis of the main message effect consists of all of the households not chosen for the survey, it is possible that the act of being surveyed itself, rather than receiving an encouragement message, explains the outcomes. As [Zwane et al. \(2011\)](#) show, the act of being surveyed may produce effects even when surveys do not contain questions specifically asking about intentions to adopt the behavior of interest. In the present context, the household survey used in the data collection lasts about 1-1.5 hours, consisting of upwards of two hundred questions that cover topics including the following: agency, empowerment, and well-being; citizen knowledge, perceptions, participation, and satisfaction; community relations; healthcare utilization and patient satisfaction; child health; and household demographics. While the survey does not contain any questions asking about the intent or likelihood of attending a community meeting, twelve questions inquire about meetings that may generally be held in the respondent's community and about the subject's beliefs regarding whether holding meetings with the LC3 Chairperson can lead to improvements in government service delivery. An additional twenty-one questions ask subjects about related topics, such as their ability to influence government service delivery, collective action in coordination with other community members, citizen responsibility, interactions with local leaders, satisfaction with the local government, and the role of the LC3 Chairperson in monitoring health facilities.

Given this topical coverage, it is plausible that the survey, rather than the messages, is responsible for the observed treatment effect. However, as [Zwane et al. \(2011\)](#) also show, survey effects dissipate over time. In this regard, note that the intervention meetings were held between roughly two and four months after the baseline survey, and that, in the time leading up to the meetings, households received the text message components of the encouragement messages. Therefore, it is likely that the treatment effect at least in part reflects an effect of the text messages.<sup>7</sup>

---

<sup>7</sup>Aside from a survey effect or message effect, yet another possibility is a Hawthorne effect. Since all subjects who received messages were given a sheet of paper to submit if they attended one of the meetings, they may have deduced that their behavior was being observed. However, given the intensity of the encouragement message interventions (possibly in conjunction with a survey effect), along with the fact that nearly thirty percent of households that attended meetings lost their submission sheets, any Hawthorne effect seems likely to be relatively minimal.



Nevertheless, any survey effect that may be present would likely operate through a similar mechanism as the encouragement messages — that is, as a result of individualized outreach. While survey effects may also act via reminder or priming mechanisms, this possibility again seems unlikely given the pervasive knowledge of the meetings created by the community-wide mobilization activities. Thus, under this interpretation, the presence of a survey effect would not pose major concern as a confounder, since both a survey effect and message effect would lead to the same conclusion: that individualized outreach induces a significant increase in public participation. Regardless of the precise explanation, the overall effect observed for the assignment to any message remains causal and fundamentally helps to frame the null effect observed for the particular message theme. The primary takeaway from this null effect is that the more granular features of encouragement messages do not significantly influence behavior.

A broader question, then, is to what extent the effect of the encouragement messages generalizes to other settings beyond the present public health setting. Given that the specific encouragement theme made little difference, it appears that the core motivation of citizens was the very opportunity to participate politically and to express their dissatisfaction with health service delivery to their elected leader. Since similar service delivery problems, all underlain by politics, span multiple sectors in countries like Uganda, it seems likely that a comparable participation intervention would also be effective even if the meetings addressed a different topic, such as education or agriculture.<sup>8</sup> Altogether, the findings demonstrate promising strategies to motivate households to provide public goods — a key priority especially in the context of political participation. With simple, low-cost messaging interventions, we may be able to substantially improve the quality of citizen participation programs while helping to organize citizens for the purpose of collective action.

---

<sup>8</sup>Indeed, it is possible that participation may be even stronger in such other contexts, as the level of spillovers associated with health could reduce the potential private gains from the meetings and thereby depress participation relative to that in other settings. However, as the underlying incentives for attending meetings in relation to any given sector are likely to be complex, it is not possible to draw any clear-cut comparative conclusions based on the currently available information.

## References

- Altmann, Steffen, and Christian Traxler. 2014. “Nudges at the Dentist.” *European Economic Review* 72:19–38.
- Ariely, Dan, Anat Bracha, and Stephan Meier. 2009. “Doing Good or Doing Well? Image Motivation and Monetary Incentives in Behaving Prosocially.” *American Economic Review* 99 (1): 544–55.
- Banerjee, Abhijit V, Rukmini Banerji, Esther Duflo, Rachel Glennerster, and Stuti Khemani. 2010. “Pitfalls of Participatory Programs: Evidence from a Randomized Evaluation in Education in India.” *American Economic Journal: Economic Policy* 2 (1): 1–30.
- Bartke, Simon, Andreas Friedl, Felix Gelhaar, and Laura Reh. 2017. “Social Comparison Nudges—Guessing the Norm Increases Charitable Giving.” *Economics Letters* 152:73–75.
- Björkman, Martina, and Jakob Svensson. 2009. “Power to the People: Evidence from a Randomized Field Experiment on Community-based Monitoring in Uganda.” *Quarterly Journal of Economics* 124 (2): 735–769.
- Brandon, Alec, John A List, Robert D Metcalfe, Michael K Price, and Florian Rundhammer. 2019. “Testing for Crowd Out in Social Nudges: Evidence from a Natural Field Experiment in the Market for Electricity.” *Proceedings of the National Academy of Sciences* 116 (12): 5293–5298.
- Casey, Katherine. 2018. “Radical Decentralization: Does Community-Driven Development Work?” *Annual Review of Economics* 10:139–163.
- Costa, Dora L, and Matthew E Kahn. 2013. “Energy Conservation “Nudges” and Environmentalist Ideology: Evidence from a Randomized Residential Electricity Field Experiment.” *Journal of the European Economic Association* 11 (3): 680–702.
- Dale, Allison, and Aaron Strauss. 2009. “Don’t Forget to Vote: Text Message Reminders as a Mobilization Tool.” *American Journal of Political Science* 53 (4): 787–804.
- Fehr, Ernst, and Klaus M Schmidt. 1999. “A Theory of Fairness, Competition, and Cooperation.” *The Quarterly Journal of Economics* 114 (3): 817–868.
- Filmer, Deon, and Lant H Pritchett. 2001. “Estimating Wealth Effects without Expenditure Data—or Tears: An Application to Educational Enrollments in States of India.” *Demography* 38 (1): 115–132.
- Frey, Bruno S, and Felix Oberholzer-Gee. 1997. “The Cost of Price Incentives: An Empirical Analysis of Motivation Crowding-Out.” *American Economic Review* 87 (4): 746–755.
- Goswami, Indranil, and Oleg Urminsky. 2016. “When Should the Ask Be a Nudge? The Effect of Default Amounts on Charitable Donations.” *Journal of Marketing Research* 53 (5): 829–846.
- Mellström, Carl, and Magnus Johannesson. 2008. “Crowding Out in Blood Donation: Was Titmuss Right?” *Journal of the European Economic Association* 6 (4): 845–863.

- Olken, Benjamin A. 2007. "Monitoring Corruption: Evidence from a Field Experiment in Indonesia." *Journal of Political Economy* 115 (2): 200–249.
- Republic of Uganda. 2000. "Laws of Uganda." Chap. 243: The Local Governments Act. Kampala, Uganda: Republic of Uganda.
- Stutzer, Alois, Lorenz Goette, and Michael Zehnder. 2011. "Active Decisions and Prosocial Behaviour: A Field Experiment on Blood Donation." *The Economic Journal* 121 (556): F476–F493.
- Thaler, Richard H, and Cass R Sunstein. 2009. *Nudge: Improving Decisions about Health, Wealth, and Happiness*. Penguin.
- World Bank. 2003. *World Development Report 2004: Making Services Work for Poor People*. Washington, DC: World Bank.
- . 2017. *World Development Report 2017: Governance and the Law*. Washington, DC: World Bank. doi:[10.1596/978-1-4648-0950-7](https://doi.org/10.1596/978-1-4648-0950-7).
- Zarghamee, Homa S, Kent D Messer, Jacob R Fooks, William D Schulze, Shang Wu, and Jubo Yan. 2017. "Nudging Charitable Giving: Three Field Experiments." *Journal of Behavioral and Experimental Economics* 66:137–149.
- Zwane, Alix Peterson, Jonathan Zinman, Eric Van Dusen, William Pariente, Clair Null, Edward Miguel, Michael Kremer, Dean S Karlan, Richard Hornbeck, Xavier Giné, et al. 2011. "Being Surveyed Can Change Later Behavior and Related Parameter Estimates." *Proceedings of the National Academy of Sciences* 108 (5): 1821–1826.

## CHAPTER 2

### Governance, Citizenship, and Accountability: A Pilot Study of People-Centered Development in the Ugandan Health Sector

#### 2.1 Introduction

The technologies needed to save millions of lives each year — mostly in low- and middle-income countries (LMICs) — have existed for decades. For example, 5.2 million children across the world continue to die annually (UN IGME 2020), mainly from causes such as pneumonia, diarrhea, and intrapartum-related complications, all of which can be prevented or treated with cost-effective solutions (Chopra et al. 2013; Liu et al. 2015). Why do governments fail to deliver available health interventions, despite their affordability?

##### 2.1.1 Local Misallocation

Although such delivery failures may be partly driven by resource misallocations at a global level,<sup>1</sup> they also stem largely from local misallocation of resources. These local misallocations may themselves span several levels, from central government expenditures to the transportation and delivery of health supplies to the organization and performance of clinical services at health facilities. The third category of “last-mile” service provision failures plays a prominent role and is rooted in widespread accountability gaps in the public health facilities of poor countries, including health worker absenteeism (Chaudhury et al. 2006), low health worker performance (Leonard and Masatu 2010), diversion of supply and medication inventories, and informal payment in the provision of services (Lewis and Pettersson 2009; McPake et al. 1999). For example, in the Ugandan public health sector, the setting of this

---

<sup>1</sup>For example, global resource misallocations may include between-country inequities and inefficiencies in public finance, economic growth, and investment, as well as market failures in the pharmaceutical industry.

study, health worker absenteeism rates exceed 47 percent ([World Bank 2016](#)). Such shortfalls constitute fundamental management problems in the healthcare systems of LMICs, leading to significant wastage of available resources.<sup>2</sup>

Strategies for improving care at local health facilities often include training health workers, providing equipment, and introducing quality protocols. However, these primarily technical interventions may be limited in impact if the same institutional structures among healthcare actors remain. The greater need may be for fundamental reform in the service delivery incentives faced by health workers and political leaders ([Pritchett and Woolcock 2004](#); [Bold and Svensson 2013](#)).

### **2.1.2 Top-Down Monitoring**

One strategy to better align the incentives for healthcare delivery is increased top-down monitoring — namely, through audits. In Indonesia, for instance, [Olken \(2007\)](#) finds that increased frequency of audits by the central government reduces corruption in village road projects. As a penalty-based mechanism to improve healthcare provider incentives, audits may be carried out by higher-level authorities such as the district health management teams that oversee local health facilities in Uganda. However, despite their promise, top-down audits also come with downsides. First, since such measures constitute a central government function, it may not be feasible for the central government to monitor the bureaucrats appointed to perform audits; consequently these individuals themselves may be corrupt ([Bardhan and Mookherjee 2006](#)). In addition, top-down approaches may require substantial resources, requiring auditors to visit a large number of facilities in their jurisdictions on a regular basis, and may therefore be difficult to sustain. Even if such audits are effective, they beget another, deeper question: what creates the political incentives to establish an auditing program in the first place?

---

<sup>2</sup>For more detailed discussion of current health sector performance in Uganda, please see [Appendix B.5.1](#).

### 2.1.3 People-Centered Approaches

As the fundamental underpinning of political systems, the preferences of the populace shape the political incentives for effective policy. Hence, the solution to ineffective governance might boil down to more effective citizenship. By the same token, interventions that instead target political incentives themselves — through organic, people-centered approaches — could lead to far-reaching reforms by altering the very political system that is responsible for quality public service provision. Indeed, in addressing the foundational elements of the political system, we may be able to produce a constellation of public service enhancements extending well beyond the establishment of a single intervention, such as auditing. In addition, by altering institutional structures, grassroots, system-driven approaches may provide a mechanism for sustaining service quality improvements in the long-run.

Given these considerations, strategies that pivot around citizens, who reside in proximity to health centers and who are capable of altering the incentives of political leaders, may promise significant impacts on healthcare delivery. Along these lines, the 2004 *World Development Report* (WDR) focuses entirely on the role of citizen participation in assuring accountability in service delivery, and the 2017 WDR centers on the essential role of governance in development ([World Bank 2003](#); [2017](#)). Despite this attention, approaches that give citizens a platform from which they can better influence government service delivery have not been widely prioritized in development policy, and relatively little rigorous evidence exists on suitable interventions.

### 2.1.4 Local Political Leaders in Uganda

In Uganda, local political leaders, who are elected democratically, have the potential to mitigate accountability problems by monitoring service provision at local government health centers. However, politicians tend to engage with the health centers on rare occasions — but have the ability to institute important changes when they do. How so? Most fundamentally, Uganda’s *Local Governments Act* invests local politicians with responsibilities that include overseeing government employee performance and government service provision within their jurisdictions ([Republic of Uganda 2000](#)). In terms of enforcing improved service

delivery based on these legal monitoring authorities, several factors make local politicians well-positioned to induce responses from health workers. First, local politicians can report negative performance to district authorities, including the District Health Officer, District Personnel Officer, Chief Administrative Officer, and District Service Commission, the body responsible for hiring health center employees. If a health worker's performance is sufficiently lacking, politicians (through their jurisdiction's council) can request a transfer of the employee. Such transfers are extremely costly to health workers, with negative consequences for their reputations and networks. According to [Raffler \(2020\)](#), about 50 percent of transfer requests are granted. Second, many local political leaders reside in offices that are proximate to the health centers, giving them easy, consistent access to the facilities. Thus, if local politicians begin to better exercise their monitoring responsibilities, this proximity should make punishment threats against health workers more credible. Third, health workers may also respond to local politicians due to the threat of social sanctions (e.g., humiliation for poor performance) within the community.

Yet, despite all of their abilities to help improve the quality of service delivery, few local leaders engage in regular monitoring activity — exemplifying the gap between the existence and the enforcement of the law that persists in many LMICs ([World Bank 2017](#)). Moreover, the constituents of local politicians exert little pressure on them to take action. Why not, given that households and communities might experience significant benefits if their democratically elected leaders adopted more measures to improve public service delivery?

### **2.1.5 Theoretical Framework**

Economic theory, in conjunction with qualitative information I have collected in the field, points to several possible explanations for the apparent institutional failures in the health sector and political system. Broadly speaking, these explanations cast such accountability deficiencies in health service delivery (*vis-à-vis* the role of political leaders) as either an incentives/information problem, a human capital problem, or both.

### **2.1.5.1 Incentives and Information**

According to the incentives and information characterization, two main constraints may lead to suboptimal politician monitoring of health service quality: (1) unobservability of politician performance by citizens and (2) incomplete information experienced by political leaders related to quality of care at local health centers. The first constraint is grounded in the principal-agent problem that characterizes the relationship between citizens and politicians. Due to the features of this problem (Ross 1973; Hölmstrom 1979), citizens (principals) may be unable to efficiently reward or sanction politicians (agents). Specifically, politician performance — with regard to the health service monitoring activities in which leaders engage — may be largely unobservable. As a result, citizens may be unable to effectively sanction politicians in elections, since they may be unsure of the actions taken by politicians and of whether politicians are to blame for poor health services. Thus, even in the context of democratic elections, citizen incomplete information may render a punishment threat to vote a politician out of office infeasible. The second-best solution for citizens would be to infer politician performance based on the quality of health service provision. However, this measure is likely to be a very noisy signal of politician performance; citizens may never be able to tease out the component of health service quality that can be attributed to politician behavior, again inhibiting them from efficiently determining rewards or sanctions for performance.

Beyond the principal-agent problem, the second constraint — incomplete information on the part of political leaders — arises from the fact that many politicians may seek healthcare elsewhere and, as a result, may be unaware of the accountability problems. Moreover, even if they seek services from the health centers, they may be treated differently and may remain unaware of the average experience of their constituents.

### **2.1.5.2 Human Capital**

According to the human capital characterization, local politicians may have inadequate training for overseeing health service delivery. Specifically, politicians may be unaware of their rights and responsibilities as elected leaders, along with effective methods for monitoring



health service delivery (e.g., gathering information on patient complaints, auditing staff attendance, spot-checking inventory stocks). Beyond methods to collect information on quality of care, politicians also need to understand how to act on the information that they gather. In this regard, they may possess inadequate knowledge of the political and institutional channels through which to pursue improvements in the healthcare system. For instance, at a local level, what strategies can politicians use to work with health center staff to integrate improved practices into facility operations? At a more centralized level, if politicians identify service shortfalls that require resolutions by higher levels of government (e.g., increased budgetary allocations, provision of additional staff), what are the proper administrative procedures that local leaders should pursue to submit requests and to advocate on behalf of their constituencies?

Observations from the field confirm that local leaders appear to be largely unaware of the actions that they can take on these fronts. Such human capital constraints likely arise, at least in part, from information asymmetries between politicians and health workers, with the latter acting as gatekeepers of important health facility information and the former being uninformed of their rights and responsibilities as elected representatives. Indeed, [Raffler \(2020\)](#) observes this dynamic in Uganda while studying local government financial accountability.

The remainder of this paper is structured as follows. In [Section 2.2](#), I describe the overall goals of this study as a pilot testing new approaches to address the health service provision gaps in Uganda. In [Section 2.3](#), I contextualize the study within the body of existing literature on governance interventions. Following, [Section 2.4](#) outlines the study's experiment design and describes in detail the study's interventions and data collection. [Section 2.5](#) presents the main results of the study, including achievements from the implementation of activities, along with the participation and engagement outcomes of the interventions. Then, [Section 2.6](#) presents an initial, condensed pre-analysis plan for a follow-on study with sufficient statistical power to enable a rigorous quantitative evaluation. Finally, [Section 2.7](#) discusses the significance of the findings, in regard to both our understanding of the political economy of healthcare delivery and the potential of the subsequent expanded study.

## 2.2 Study Objectives

Carried out in collaboration with the Office of the Prime Minister (OPM) and a non-profit organization, Progressive Health Partnership (PHP), in Uganda, this study consists of a feasibility assessment and pilot randomized controlled trial of the following two governance interventions: (a) increased citizen participation through quarterly feedback, reporting, and accountability meetings with local politicians to discuss health service quality and (b) skills training for politicians on monitoring local government health centers. Designed based on the theoretical framework described in Section 2.1.5, the OPM’s lessons learned from its accountability programs, and PHP’s experiences in its partner communities, the interventions have the potential both to produce significant impacts on health service delivery and to be implemented at greater scale through the public sector.

In preparation for a longer-term study with greater statistical power, the primary objective of this pilot has been twofold: to evaluate the feasibility of (1) the interventions and (2) the data collection activities. Through this study, I examine the nature and level of local responsiveness to the interventions by both citizens and political leaders, the feasibility of scaling up the programs, and the qualitative effects and mechanisms of the interventions. Equally important, I evaluate the feasibility of the data collection activities — a key objective given the potential political sensitivities of the survey questions.

For the OPM, this small-scale study has provided an opportunity to test new accountability strategies intended to build on its Baraza Program, a local citizen advocacy initiative that the OPM has been carrying out since 2009. Based on the results of this pilot, the OPM is committed to collaborating on a larger study. Altogether, the pilot serves to provide critical evidence for proof of concept, establishing the framework for a rigorous evaluation.

## 2.3 Literature Review

Only a few other randomized evaluations of similarly motivated interventions have been performed. In regard to community organization, [Björkman and Svensson \(2009\)](#) find significant impacts of a participatory intervention in the Ugandan health sector that mobilizes

residents to establish health provider monitoring activities implemented and sustained by the community. In encouraging citizen themselves to adopt public sector monitoring practices, this approach centers on the role of the citizen-provider relationship. In contrast, the present study hinges on the relationship between citizens and politicians, the fundamental underpinning of any democratic system of governance. In this context, changes initiated in response to citizen participation should act through improved performance of existing monitoring responsibilities by politicians — not through the establishment of new, parallel monitoring institutions.

Studying corruption, [Olken \(2007\)](#) observes little effect of an “accountability meetings” intervention that again focuses on the citizen-provider rather than the citizen-politician relationship. Because public service providers do not face direct electoral punishment threats from citizens, the citizen-provider approach may be less effective. Pointing to the potential impacts of the meetings intervention in this study, [Bidwell, Casey, and Glennerster \(2020\)](#) find that political information — conveyed through public debates — affects voting behavior, leading to campaign expenditure responses by candidates and greater financial accountability of subsequently elected leaders.

In regard to politician training, [Raffler \(2020\)](#) evaluates the impact of a similar local political leader training and information program in Uganda, which leads to improved political oversight of bureaucrats involved in managing finances and local development projects. [Ravanilla \(2016\)](#) finds that a training workshop combined with performance-based rewards in the Philippines leads more qualified individuals to serve in public office. The present study aims to build on this previous work by not only examining intermediary outcomes such as politician behavior but also welfare impacts on health. I also examine aspects of job training that have not yet been rigorously tested.

## 2.4 Methods

### 2.4.1 Experiment Design

Along with the OPM, I collaborated with PHP to implement the interventions in a stratified, cross-randomized,  $2 \times 2$  design across twenty subcounties/town councils/divisions throughout the Ankole Region of Uganda. As equivalent administrative units across rural, periurban, and urban areas, respectively, each subcounty, town council, and division has a government health center referred to as a Health Center III and is headed by a democratically elected council chairperson. To account for the potential influence of the underlying political environment on intervention effects, I used data from the Electoral Commission of Uganda to stratify the experimental randomization by the margin of victory of council chairpersons in the recent 2016 election. The margin of victory may be a proxy for the level of political competitiveness and possibly for the quality of democratic functioning in each locality. Specifically, I divided the population of subcounties/town councils/divisions into five strata: the first four strata matched the quartiles of the margin of victory variable, while the fifth stratum included all of the localities in which the chairperson ran for office unopposed in the recent election.<sup>3</sup>

FIGURE 2.1: EXPERIMENT RANDOMIZATION MATRIX

		T2: Politician Training	
		NO	YES
T1: Citizen-Politician Meetings	NO	Pure Control (5 localities)	Training Only (5 localities)
	YES	Meetings Only (5 localities)	Interaction (5 localities)

*Notes:* This figure depicts the study's cross-randomized design. The twenty sample localities were evenly divided amongst the four study arms. The interaction arm included both the training and meetings interventions.

Within each of the five strata, I randomly selected four localities, which I then randomized

<sup>3</sup>For more information on the study population of localities, please see Appendix B.1.

across four groups: pure control, citizen-politician meetings only, politician training only, and interaction. These four arms, encompassing both single and joint implementation of the interventions, are designed to provide insight into key constraints on health service provision, based on the theoretical framework presented above. Although the scale of the pilot does not provide sufficient statistical power for rigorous quantitative analysis of the intervention effects, I implemented the stratified, randomized design to inform qualitative observations and to establish a replicable framework for a larger study.

## **2.4.2 Intervention Design and Mechanisms**

### **2.4.2.1 Citizen-Politician Meetings Program**

In the meetings treatment arm, PHP organized quarterly meetings in each locality between the Council Chairperson and citizens to discuss local healthcare quality, creating opportunities for regular feedback and reporting. Within each administrative unit, PHP carried out two meetings — one in the parish/ward that contains the Health Center III and another in a randomly chosen parish/ward.<sup>4</sup> The first location helped to ensure that the most likely users of the health facility participated in a meeting, while the second location served to increase intervention coverage and to make the meetings as accessible and as decentralized as possible. Although PHP originally planned to carry out the intervention over the course of three quarters, the program implementation was paused due to the Covid pandemic after eight of the ten meeting arm localities had been covered in the first round of activities.<sup>5</sup> This round took place between January and March 2020.

To foster widespread, representative participation, PHP implemented a systematic and multipronged mobilization procedure, which primarily consisted of making announcements throughout communities with a mobile public address system, hanging announcement posters, sending text message reminders to local leaders, and working with community health workers, village leaders, and other local influencers to advertise each meeting in advance to the population-at-large. To ensure effectively run sessions, trained facilitators mediated the

---

<sup>4</sup>Parishes are lower-level administrative units within subcounties, while wards are lower-level administrative units within town councils and divisions.

<sup>5</sup>The organization and the research team are currently assessing the possible resumption of the study activities.

meetings according to standardized guidelines. In brief, each meeting followed a standard agenda, consisting of the following: (1) introductions and establishment of general conduct expectations, (2) small-group discussions along with optional submission of anonymous written comments, (3) open discussion, including a formal response to community concerns from the LC3 Chairperson, (4) presentation of an action plan by the LC3 Chairperson and closing remarks. Given potential political sensitivities, the meeting implementation guidelines emphasized the importance of maintaining neutrality and remaining focused on constructive steps to improve healthcare delivery.

In the context of the theoretical framework outlined in Section 2.1.5, the citizen-politician meetings intervention aims to address the unobservability problem of the principal-agent relationship, by giving citizens an opportunity to adopt measures to directly observe politician performance. Specifically, the meetings intervention enables citizens to hold political leaders more accountable by gathering regular (quarterly) information on politician behavior vis-à-vis local health facility service delivery. Such political information pathways have been shown to be influential by a range of other studies.<sup>6,7</sup> Findings from my qualitative research also confirm the plausibility of this pathway, as do focus groups in Uganda conducted by Larreguy et al. (2017). Overall, targeting the unobservability of the citizen-politician relationship through community meetings offers a promising approach, since the second-best solution to the principal-agent problem — inferring politician performance based on the quality of health service provision — is likely infeasible. Beyond this aim, the meetings intervention also serves to address the second constraint discussed in Section 2.1.5.1 regarding politician incomplete information on health service quality. Because citizens — from their own experiences with local health facilities — appear to have sufficient knowledge of service quality, the meetings provide an opportunity for them to convey this information to political leaders.<sup>8</sup>

---

<sup>6</sup>See Pande (2011) for a review. For more recent studies, see, for example, Grossman and Michelitch (2018); Bidwell, Casey, and Glennerster (2020), Gottlieb (2016); Paler (2013); and Fujiwara and Wantchekon (2013).

<sup>7</sup>In the present study, however, it is possible that the meetings may also serve to increase citizen motivation. Disentangling motivation and information effects is a challenge in the literature and would be a suitable topic for future research on the meetings intervention. Additionally, it is possible that the meetings not only may provide citizens a source of information on politician performance but also may help to overcome citizen incomplete information regarding political and health rights.

<sup>8</sup>For this channel, note that the role of citizen information on health service provision is not impaired by

Altogether, these meeting intervention mechanisms suggest that greater citizen-politician interaction could help overcome key barriers to health service quality. Given the potential returns to citizen-politician meetings, what keeps citizens from organizing such meetings in the status quo? Several additional constraints may explain limited political participation by citizens, including the following: (1) organization, (2) information, (3) transaction costs, (4) free-rider effects, and (5) powerlessness. First, citizens likely require robust leadership to help provide structure for collective organization and to offer them a concrete course of action. Without such leadership, citizens may be unlikely to participate. In a perfectly functioning democracy — with perfect information flow between citizens and politicians and complete accountability of political leaders — we might expect the politicians themselves to provide the necessary leadership and resources to organize meetings with their constituents. In imperfect democracies or other political systems, third-party intervention may be required to invigorate collective action. Second, in addition to these organizational constraints, information deficiencies may limit citizen participation. Informational gaps may be as basic as a lack of awareness regarding opportunities to participate in collective action, which citizens may need for the purposes of advance planning. Citizens may also experience incomplete information regarding their political rights, as well as their right to free, quality public healthcare in Uganda. Even if citizens are aware of their health and political rights, they may experience incomplete information regarding the political channels and systems through which to exercise them. Third, the need for transportation may create a transaction cost that amounts to a physical barrier to citizen participation. Especially in large jurisdictions where major political events are typically held in the center, citizens from more distant areas may be inhibited from participating.

Fourth, since democratic participation is a public good, attendance at such citizen-politician meetings is subject to free-rider effects that discourage participation. Fifth, since individual citizens (particularly non-elite citizens) in the political system are akin to price-takers in a competitive goods market, the complaints of one citizen alone are unlikely to

---

the observation that information on health service quality is a noisy signal of politician performance in the second-best solution to the principal-agent problem. In the principal-agent setting, the noisiness pertains to the ability of citizens to *attribute* observed health outcomes to politicians. In contrast, in the politician incomplete information mechanism, citizens only need to be generally aware of health service quality, in order to submit complaints to their leaders that spur politician action.

influence the behavior of a political leader or create sufficient politician perceptions of electoral threats. This dynamic amounts to low bargaining power on the part of the average citizen and may be expressed by citizens as a sense of powerlessness — either real or perceived — to influence politician behavior. Taken together, these last two constraints of free-riding and price-taking make collective action an essential element of political processes.<sup>9</sup>

Providing a platform for citizens to engage with their local leaders, the meetings intervention aims to overcome the foregoing constraints. In particular, the community accountability meetings are organized by a non-governmental organization, thereby eliminating leadership and organizational constraints. In addition, the implementation of the meetings is accompanied by substantial mobilization efforts, which serve to alleviate informational barriers to participation. The meetings also take place in two locations within each governing jurisdiction, helping to make the events as accessible as possible. Finally, by bringing many community members together for the common cause of political accountability, the meetings address the free-rider and powerlessness constraints.<sup>10</sup> Of note, many of the constraints to political participation may disproportionately affect poorer citizens, a group that may be more likely to seek care from government health facilities. If wealthier citizens are less likely to obtain services from the public health centers and can also more easily access political leaders, politicians may develop an inaccurate viewpoint of local healthcare quality. Hence, bringing politicians and the general population together may help to ensure more accurate representation.<sup>11</sup>

---

<sup>9</sup>Yet another constraint that may limit participation is motivation. Even when citizens have all of the necessary information and resources to adopt an accessible course of action, they may need additional persuasion to inspire them to join a collective action effort. For more discussion of motivation-related constraints, please see Chapter 1 of this dissertation.

<sup>10</sup>One possible constraint on the effectiveness of the accountability meetings is the existence of social networks linking politicians and health workers. However, the inherent aim of the meetings is to generate sufficient citizen pressure through mass action to overcome any such opposing pressures on politicians. The influence of social networks on politicians is also another topic to address in a larger-scale study.

<sup>11</sup>For additional discussion of how the meetings intervention may impact health service provision, particularly in the context of a political environment with dysfunctional electoral processes, please see Appendix B.4.



### 2.4.2.2 Politician Training Program

In the skills training treatment arm, PHP followed a standardized curriculum to train chairpersons on their rights and responsibilities and on methods for monitoring local health facilities and taking action based on their findings. The curriculum addressed topics such as identifying patient complaints, checking for the presence of health workers, monitoring inventories, instituting improvements in local management practices, and using proper legal channels to pursue changes that may require intervention at higher government levels. In the context of the theoretical framework outlined in Section 2.1.5, the training intervention aims to address the human capital constraints that political leaders may experience.<sup>12</sup> As with the meetings program, PHP originally planned to carry out the training program over the course of three quarters, with a primary training workshop followed by two refresher workshops. However, because of the Covid pandemic, the organization thus far has only been able to complete the first round of training, which took place in January 2020 prior the launch of the meetings program.

### 2.4.2.3 Impact Pathways

The programs may produce impacts in two primary areas: health and citizen satisfaction. To produce an impact on health outcomes, each intervention must induce a relatively complex cascade of effects on several intermediary outcomes. Specifically, the programs must first alter political incentives and dynamics. For the meetings intervention, such political effects must begin with strong grassroots participation (also referred to as a ‘first-stage’ effect), which itself must then be successful in influencing politician behavior. For the training intervention, the program must produce significant learning outcomes, which then also must lead to changes in politician behavior. Under both programs, the effects on political dynamics must in turn affect operational practices at health facilities. Finally, the changes in health center operations must then translate into improved health outcomes. At any stage in this pathway, binding constraints external to the interventions — for example, capacity lim-

---

<sup>12</sup>As another possible channel for this study arm, the training may not impart any new skills to politicians but may merely serve to remind them about health service delivery shortfalls, leading them to prioritize this area for oversight. As the pilot study is not able to tease apart these two interpretations, this distinction would be a priority for future research.

itations at health facilities — may thwart the realization of an impact, even if the employed programmatic strategies are effective.

Health outcomes, of course, are not the only contributor to social welfare. By fostering democratic participation, the meetings program in particular may also impact household and community well-being in areas related to self-determination, agency, and empowerment. Such impacts would likely be more intrinsic to the process of program implementation and therefore, compared to potential health impacts, would not require such an involved cascade of effects.

### **2.4.3 Data Collection**

Carried out by PHP’s field interview team, the data collection for the study consisted of a household survey, politician survey, and health facility quality assessment across all study arms, along with a meeting survey in the citizen-politician meetings arm. While both pre- and post-intervention data collection was originally planned for the household, politician, and health facility surveys, the post-intervention data collection has not yet occurred due to the aforementioned pause in the project activities as a result of the Covid pandemic. However, because the primary objective of the study was feasibility assessment rather than impact evaluation, the round of baseline data collection, combined with nearly one full round of the midline meetings survey, proved sufficient for satisfying the core research goals. All baseline data collection took place between October and November 2019, while the meetings survey occurred between January and March 2020 in conjunction with program implementation.

All data collection instruments were electronically programmed via SurveyCTO and passed the rigor of multiple levels of development and review, including pre-testing with live subjects, quality assurance mechanisms for the fidelity of language translation, and bench testing for programming quality and errors. Each data collection tool was accompanied by an extensive training manual that covered general conduct, interviewer guidelines, data collection procedures, and question-specific instructions and definitions. In addition, for each survey tool, field interviewers went through a tailored, competitive training workshop of 5-8 days in length that concluded with a written examination. Due to the relatively small size of the study and accompanying budget, audits of the data collection process were conducted

by the manager and assistant manager of the data collection team. The research received ethical approval from the University of Michigan Institutional Review Board, the Makerere University Research Ethics Committee, and the Uganda National Council for Science and Technology.

#### **2.4.3.1 Outcomes of Interest**

Based on the health impact pathway described in Section 2.4.2.3, the data collection for this study is designed to measure political outcomes, health services outcomes, and individual-level health impacts. Political outcomes — which demonstrate intervention success in altering local political dynamics — include participation in meetings with local leaders, citizen perceptions of politicians, perceptions of health center quality by politicians, and changes in job practices by local leaders. Health services outcomes — which demonstrate the extent to which political changes alter facility practices — include a collection of measures that fall into one of four categories: (1) healthcare utilization, (2) healthcare quality, (3) intervention delivery and adoption, and (4) patient satisfaction. Health impacts include child mortality, weight-for-age, and mid-upper arm circumference.<sup>13</sup> In addition, apart from the health impact pathway, the data collection includes measures of citizen satisfaction to capture possible impacts stemming from enhanced self-determination, agency, and empowerment.

For the purpose of the present pilot study, with limited statistical power to evaluate all of the foregoing outcomes, I primarily focus on demonstrating the ‘first-stage’ effects of the interventions. Toward this end, key outcomes include measures of politician and citizen participation in the meetings, such as attendance, gender inclusiveness, elite vs. non-elite participation, and problems reported; feedback from households regarding their participation in the meetings; measures of politician participation in the training program; and feedback from political leaders regarding the quality of the training program.

In a longer-term study with greater statistical power, the aforementioned data collection will enable not only an evaluation of impact but also an analysis of which constraints on health service provision appear to be most active in the intervention mechanisms. In such

---

<sup>13</sup>For more detailed discussion of the health outcomes of interest, focusing in particular on current health-care delivery gaps in Uganda and the potential intervention impact pathways, please see Appendix B.5.2.

a study, I will also aim to differentiate between alternative possible impact pathways that could be fostered by the interventions.

#### **2.4.3.2 Household Survey**

The household survey randomly sampled households from villages across the control and treatment localities. The survey lasted 1-1.5 hours and covered topics including following: agency, empowerment, and well-being; citizen knowledge, perceptions, participation, and satisfaction; community relations; healthcare utilization and patient satisfaction; child health, including anthropometric measurements; and household demographics. To define the catchment area for the survey, I carried out the following procedure. First, recall that PHP conducted the meetings intervention in two locations within each of the meeting treatment localities. One meeting took place in the parish/ward containing the Health Center III, while the other took place in another randomly chosen parish/ward. These two locations within each locality formed the basis of the household survey sampling. However, because the meeting implementation locations only applied to the meetings treatment group, the pure control and politician training only groups required an equivalent design to guide the household survey sampling process. Therefore, in such non-meeting localities, the data collection team similarly identified the parish/ward containing the Health Center III and randomly selected another parish/ward.

Following, for each selected parish/ward location, the data collection team identified a cluster of three central villages in which the survey sampling would take place. For the parish/ward containing the Health Center III, the first of the three villages was defined as the village containing the Health Center III. The second and third villages chosen were the nearest neighbors to the first village. For the other parish/ward that did not include the Health Center III, the first of three villages was defined as the village containing the parish/ward administrative headquarters, while the remaining two villages were again the nearest neighbors. Appendix B.2 contains the determination criteria for the nearest neighboring villages. At the conclusion of the village selection process, each main study locality contained two clusters of three villages each, which formed the sampling area for the household survey.

The sampling procedure then proceeded according to the urbanization level of each lo-

cality. For the three of the twenty localities in the sample that were more densely populated town councils or divisions, the field interview team employed a random route method to sample households, as pre-survey enumeration of all of the resident households was impractical due to budgetary constraints. In accordance with this procedure, interviewers began at the center of each three-village cluster (either the Health Center III or the local government ward headquarters) and then followed instructions that led them on a random route to sample households, while always remaining within the boundaries of the three villages. Immediately after sampling, field interviewers scheduled the survey administration with the household.

For the other seventeen, less densely populated localities, PHP's data collection team, in advance of the baseline household survey, carried out enumeration of all households in the sampling area to form the sampling frame. To complete the household enumeration process, the data collection team visited the village chairperson in each of the sample villages to request a copy of the village's household registry. Members of the data collection team then reviewed the registry with the chairperson, verifying that the list of households was comprehensive, accurate, and up-to-date. To be eligible for the survey, households were required to have one or more children under five as well as a respondent above 18 years of age. Therefore, as an initial eligibility screen, the data collection team asked the chairperson whether each household contained one or more children under five. In general, village chairpersons are knowledgeable about the households in their community. However, the question responses included a "Don't know" option. All households with "Yes" or "Don't know" responses for the child under five question were included in the sampling frame. In all cases, eligibility was confirmed directly with the household at the time of survey administration.

From the sampling frame, I then randomly selected households for the baseline survey, along with a list of replacements in case of ineligible, untraceable, or non-participating households. If an interviewer visited a household for the survey and determined that the respondent and/or household was temporarily absent, the household would be revisited. If a household was visited three times and an eligible and competent respondent remained unavailable each time, the household was dropped from the sample and replaced. To meet the sample goal of 500 households, twenty-five households from each locality were initially selected for the sample. To ensure that each cluster of sample villages with the localities

would be roughly equally represented in the sample, the twenty-five households were divided into thirteen households from each Health Center III village cluster and twelve households from each non-Health Center III village cluster. Ultimately, thanks to the productivity of the field interview team, the survey sample included 553 households.

#### **2.4.3.3 Politician Survey**

The politician survey lasted 1-1.5 hours and covered topics including the following: job background, knowledge, perceptions, and practices; health sector knowledge and perceptions; healthcare utilization; job-related activities in the health sector; constituent perceptions and interactions; and community relations. The sample for the survey was pre-defined, consisting of all of the Council Chairpersons for the twenty administrative units included in the study. Securing interview appointments with these local political leaders proved to be a relatively difficult process. In the end, the final survey sample consisted of sixteen of the twenty chairpersons, as the remaining subjects either refused to participate, could not be contacted, or were deceased.

#### **2.4.3.4 Health Facility Quality Assessment**

The health facility quality assessment consisted of an unannounced audit visit along with a more extended questionnaire, which was scheduled at the time of the initial audit visit. For the unannounced visit, the interviewer arrived at the health facility prior to the official opening time of 8:00AM and then discretely recorded observational data regarding the facility's operations, including the actual opening time. If and when the health center opened, the interviewer then conducted a personnel audit by comparing the facility's official personnel roster with the actual staff attendance. The scheduled questionnaire portion of the quality assessment lasted 2-3 hours and covered topics including the following: service delivery, patient experience, monitoring and supervision, community relations, infrastructure, equipment, drugs and supplies, healthcare utilization, and use of locally allocated healthcare funds. The equipment-related questions involved physical verification of the availability and functionality of selected items. Similarly, questions about inventories focused on a list of tracer medicine and supplies and included physical verification of the available units, along

with collection of administrative stock records. As with the politician survey, the sample for the health facility quality assessment was pre-defined, consisting of all of the Health Center III's in the twenty administrative units included in the study. As all of the health facilities consented to participate in the survey, the final sample consisted of twenty health centers.

#### **2.4.3.5 Meeting Survey**

The meeting survey involved a data collector-observer who recorded information about each community meeting as it was conducted, covering the following topics: attendance figures, according to different sub-populations (e.g., male vs. female, elite vs. non-elite); participation levels, again according to different sub-populations; and meeting discussion content.

## **2.5 Results**

Both of the interventions yielded important results, demonstrating the promise of a follow-on study with greater statistical power. The key results fall into two categories: (1) achievements from program implementation and (2) outcomes of the interventions.

### **2.5.1 Implementation Outputs**

From the standpoint of implementation, the pilot study has been a major success. In particular, the pilot has provided an opportunity to develop key program and research materials and protocols that will lay the foundation for an expanded study in the future. In the course of developing these materials, our research team adhered to stringent quality assurance and quality control measures — based on the idea that if the project sought to improve institutional quality and accountability, then we must exemplify those same priorities ourselves. Toward this end, implementation materials have included a wide range of standard operating procedures and reporting systems to guide each activity. This section highlights some of the most critical outputs.

### 2.5.1.1 Research Implementation

As described above, research materials produced under the pilot have included extensive household, politician, health facility, and meeting surveys, all programmed electronically and accompanied by comprehensive training manuals, workshops, and examinations. In addition, our research team has developed in-field data auditing tools, along with a large body of computer code to facilitate the sampling process, carry out regular data quality checks, and perform final data processing. We have also created various reporting templates for the data collection team, as well as comprehensive checklists for all major activities, such as interviewer training and fieldwork, to ensure efficient implementation of the research.

In practice, the pilot has provided strong proof of concept for the feasibility of the data collection activities. Despite the political sensitivities of the research, the data collection exceeded expectations with the relative ease of its implementation. The household survey had a refusal rate of only 1.3 percent, and respondents expressed few reservations about the questions asked of them. While the respondents for the politician survey were somewhat challenging to secure for an appointment, many of them became much more enthusiastic about their participation once they learned further about the topics that the questionnaire addressed. The lessons learned from working with this group of political leaders to foster their participation in the research will be especially valuable for future data collection. The health facility quality assessment also initially posed concerns for our research team, in light of both its level of depth and its focus on health worker performance. However, as all health centers participated in the questionnaire, the survey was a significant success. To avoid disrupting the normal work flow of health facilities, our team developed approaches to adapt to the facility activities in carrying out the data collection. Ultimately, the questionnaire content proved to be well tailored to Health Center III's, capturing important information about their operations. For all of the foregoing surveys, our team of field interviewers also reported that they felt respondents answered questions with relative honesty, an encouraging indicator of the quality of participation that we were able to secure. In addition, we experienced no adverse events in the course of any of the data collection. Lastly, the meeting survey presented its own challenges given the difficulties of collecting a wide breadth of ob-



servational data in real-time. However, with logistical tools developed by our research team and a clear plan of action for the conduct of the meetings, the survey progressed smoothly. Altogether, the quality of the data collection instruments, combined with the corresponding implementation protocols that we have created, will help enable an efficient transition to an expanded study.

### **2.5.1.2 Program Implementation**

For the implementation of the two interventions, our team developed a similar breadth of tools. First, for the meetings intervention, we developed a handbook for meeting facilitators entitled *Citizen-Chairperson Dialogue Meetings: Implementation Guidelines*. The manual reviews the objectives of the meetings and the roles of meeting facilitators, presents background information on local government and health facility operations, establishes guidelines for the meeting mobilization process, and describes the process of meeting implementation, covering topics such as the logistics, discussion content, and meeting management.

For the training intervention, we developed two additional handbooks to guide its implementation. First, for the chairperson-trainees, we developed the *Health Sector Performance Initiative Health Leadership Manual*, available in both English and the local language of Runyankore. This manual for trainees presents a curriculum covering the LC3 Chairperson's roles and responsibilities in the government, detailed background information on health sector operations and related government stakeholders, techniques for monitoring local health service quality, and guidelines for taking action to improve healthcare delivery. Accompanying the trainee manual, our team developed the *Health Sector Performance Initiative Training Guide* for the facilitators of the training workshop. The training facilitation manual provides an extended series of educational activities to carry out in presenting the contents of the trainee handbook at the training workshop. The facilitation manual is also accompanied by slides for the presentation of all of the material.

Alongside the development of program implementation guidelines, we also established a system of 'activity-based reporting,' through which we designed reporting requirements around each component activity of the interventions. Each report template contains detailed instructions with step-by-step procedures for the corresponding activity, a main report body,

and an item-by-item checklist to guide the preparation for and implementation of the activity. By documenting and accounting for the activities so extensively, we minimized the error rate over the course of implementation. The checklists, which cover many of the mundane and easily forgettable tasks involved in program implementation, also served to free up the energy of the implementation team to focus on more intangible or discretionary tasks that also contribute to program quality.

Altogether, the collection of program guidelines and reporting tools help to ensure standardized, efficient, and high-quality implementation of the activities across different localities and facilitators. The protocols also make the interventions highly replicable, which will help to facilitate a larger-scale study subsequent to this pilot. Moving forward, in preparation for scale-up and based on lessons learned, our team will make additional tweaks to all of the implementation materials to further fine-tune them to the program activities.

## **2.5.2 Program Outcomes**

### **2.5.2.1 Politician Training**

The training intervention brought together the LC3 chairpersons from the ten training localities for a three-day workshop that covered the curricular materials described in Section [2.5.1.2](#). Initially, our program team was uncertain of what to expect from the training; in fact, we could not even be confident that the political leaders would attend the workshop. However, in terms of both participation and learning outcomes, the program recorded achievements that exceeded expectations.

In particular, as the workshop drew the trainees away from their day jobs, and as a majority of the politician-trainees had to travel a significant distance to the training venue, the attendance of nine of ten chairpersons from the training treatment arm was itself a significant achievement. The tenth chairperson who did not attend sent another local leader as a delegate. While trainees sometimes circulated in and out of the workshop activities, we always achieved a participation rate of at least eighty percent throughout the workshop. More concretely, the training proved constructive on two main levels: (1) knowledge acquisition by the local leaders and (2) cross-community exchange of local government and health sector

experiences. First, while the politician-trainees had some background knowledge on their general roles and responsibilities, the training filled critical knowledge gaps related to the operations of the health sector, the roles of political leaders in monitoring service provision, and remedies for performance problems in healthcare delivery. Many attendees remarked that they knew that they were capable of playing key roles to improve life in their communities, but that no other programs had empowered them with the knowledge and training to do so. Based on the comments from the trainees, it is clear that LC3 chairpersons occupy a unique but neglected position in their communities from which they can help foster more accountable local governance and service delivery. Second, in addition to the core training objectives, the workshop fostered a rich exchange of experiences between all of the chairpersons in attendance. By sharing lessons from their own communities, the trainees had a valuable opportunity to learn from one another, complementing the workshop curriculum.

At the conclusion of the training, the facilitators solicited feedback from the politician-trainees regarding the workshop. All trainees offered overwhelmingly positive comments, on topics ranging from the pedagogical methods employed to the content presented. They expressed significant gratitude for the opportunity to participate, along with enthusiasm to return for future workshops. In addition, they suggested that the government ultimately integrate such workshops into standard orientation activities for political leaders. Altogether, the observations that our program team made during the training and the feedback that we received afterward evinced the value of the activities. While thus far, due to the Covid pandemic, it has only been possible to carry out one of three rounds of the politician training intervention, the achievements to date have been sufficient to demonstrate the significant promise that this program holds to impact health service provision.

### **2.5.2.2 Community Accountability Meetings**

The meetings intervention proved similarly successful, achieving strong stakeholder engagement. Table 2.1 displays summary statistics on meeting attendance and participation. The data come from a sample of sixteen meetings across eight localities, due to the Covid-related pause in the intervention implementation before the two remaining localities in the meetings treatment arm could be covered. Based on population figures from the baseline household

survey sampling frame, an average of 19.7 percent of households in each local catchment area sent one or more members to the meetings, with this measure of participation ranging from 12.5 percent to 33.0 percent across the meeting locations. On average, meetings attracted more than one hundred attendees, with the vast majority of these participants being members of the general community as opposed to individuals with official government leadership positions. In addition, these community members primarily came from the non-elite portion of the population, with members of the elite accounting for an average of only 1.2 percent of attendance by general community members.

TABLE 2.1: MEETING ATTENDANCE AND PARTICIPATION SUMMARY STATISTICS

	Mean
Proportion Catchment Area Households Attended	0.197
Total Attendance	109.688
Proportion General Community Members	0.956
Proportion Politicians	0.039
Proportion Civil Servants	0.004
Total General Community Member Attendance	105.125
Proportion Male	0.372
Proportion Female	0.628
Proportion Elite	0.012
Total Comments by General Community Members	123.188
Proportion Male	0.389
Proportion Female	0.611
Proportion Elite	0.037

*Notes:* This table displays summary statistics for the meetings intervention. The main attendance figures are based on the number of individuals present, as opposed to the number of households represented. Each mean is an average across the sixteen meetings in the sample.

In terms of gender representation, an average of 62.8 percent of the general community members present at each meeting were women. While the program aimed to elicit participation from typically underrepresented groups like women, it is also possible that the preferences of women — who in Uganda occupy the traditional role of managing household health matters, and who more frequently come into contact with the local health centers — make them more likely to attend the meetings. The discussions at the meetings reflected the composition of community members present, with men and women contributing meeting comments at rates commensurate with their level of representation. While elite community

members appear to make comments at a higher rate than their level of representation, only 3.7 percent of comments on average are attributed to elites. This observation implies that the meetings do not suffer from elite capture, a major concern in local government affairs.

Qualitatively, the discussions at the meetings were vibrant, with ordinary citizens openly and passionately voicing their complaints about health service delivery. The most common complaints included health worker absenteeism and shortage, limited working days/hours of the local health center, lack of drugs and supplies, charges or bribes for services, and disrespect by health workers. In regard to the meetings themselves, community members expressed significant appreciation for the activities, indicating that it was uncommon to hold such gatherings and to be given a platform from which to express their concerns.

TABLE 2.2: PREDICTORS OF MEETING ATTENDANCE BY HOUSEHOLDS

	Attended Meeting
Satisfaction with Health	0.000 (0.029)
Satisfaction with Community	-0.074** (0.029)
Able to Influence Local Government	0.140* (0.081)
Able to Improve Government Health Services in Community	-0.028 (0.091)
Attended Village Meetings	0.005 (0.072)
Attended LCIII Meetings	-0.236** (0.095)
Community Would Join Together to Improve Health Services	-0.088 (0.083)
Chances of Community Pressuring Chairperson	-0.029 (0.032)
Have Sufficient Information to Evaluate Local Health Services	-0.068 (0.075)
Drugs and Supplies Usually Available	0.096 (0.080)

Health Center Staff Normally Attend Work	0.109 (0.119)
Travel Time to Health Center	-0.000 (0.001)
Number Health Center Visits in Past 12 Months	-0.002 (0.010)
Number Visits to Other Providers in Past 12 Months	-0.009 (0.011)
Number Household Members	-0.013 (0.015)
Respondent Highest Grade in School	0.016 (0.012)
Household Head Highest Grade in School	-0.015 (0.012)
Economic Index	0.009 (0.022)
Mean Dependent Variable	0.416
Observations	231

*Notes:* This table displays the results of the meeting attendance predictors analysis. Heteroskedasticity-robust standard errors are shown in parentheses below the regression coefficient estimates. \*Significant at 10% level, \*\*Significant at 5% level, \*\*\*Significant at 1% level.

Alongside the observational data from the meetings, I also use the baseline household survey data linked with attendance data to assess the predictors of household attendance at the meetings. Table 2.2 presents the results of a linear probability model regressing meeting attendance on a range of baseline dimensions reported by households. Overall, relatively few covariates are significant predictors of household attendance. First, households that report greater satisfaction with the community are, for each level of satisfaction measured on a 1-5 Likert scale, 7.4 percentage points less likely to attend meetings. Second, households that report that households like their own are able to influence the decisions made by local government officials in the community are 14.0 percentage points more likely to attend meetings. Third, households that report having sent a member to one or more Local Council III meetings over the past 12 months are 23.6 percentage points less likely to attend the

accountability meetings. As the Local Council III is the top governing body in each locality and often attracts the participation of elites, this large effect may reflect an aversion of local elites to attend the meetings, perhaps because they perceive them (correctly) as events for the general community. Of note, education and economic status are not significant predictors of meeting attendance. These findings suggest that households from a range of backgrounds were attracted to the meetings — an encouraging result in light of the program’s goals related to representation.

Aside from attendance by households, the meetings also achieved strong participation by the chairpersons. In general, chairpersons found the meetings to provide an appealing opportunity to meet with their constituents. However, in two of the sample localities, the chairpersons declined to participate, providing explanations that insinuated their expectation of a monetary inducement for attending the meetings. In a third locality, the chairperson had to forego one of the two meetings due to an emergency. Although the sample size is small, these outcomes amount to an intervention adoption rate by chairpersons of roughly 70 percent. I interpret this adoption rate as strong, even though chairperson attendance would ideally be perfect, given the central role of the chairpersons in the program. In addition, even in the absence of chairpersons, meetings remained at least partially productive. In all of the instances in which the chairpersons did not attend, they delegated the vice chairpersons to attend the meetings in their place. Meetings involving the vice chairpersons were more challenging, as the deputies had less authority and also often less knowledge compared to the chairpersons. Nevertheless, our program team felt that these meetings were still helpful in bringing community members together to address important health matters. Furthermore, while it has not been possible to observe the effects of repeated meetings due to the Covid-related activity stoppage, the program team felt that the non-participating chairpersons likely would have become more cooperative as more meetings were implemented. The team additionally reported that chairpersons who were also part of the study’s training arm tended to be more cooperative, suggesting complementarity between the meeting and training programs.<sup>14</sup> Moving forward, our program team will explore further strategies to

---

<sup>14</sup>This observation may be due to the lessons that the chairpersons learned during the training, but it may also be due to the prior relationship that our staff had established with them as a result of the training — or a combination of both.

induce maximal attendance by chairpersons.<sup>15</sup>

Altogether, the varied findings on the meetings intervention demonstrate the program's success in terms of both level of attendance and level of representation. In line with the original program objectives, the meetings induced significant participation from a diverse group of households. As with the training program, the outcomes provide strong evidence in support of the meeting program's potential to ultimately impact health. In addition, in communities where the LC3 chairperson was also among the group of leaders who received the training intervention, our program implementation team reported that the meetings seemed more effective, with the chairpersons suggesting more specific and more practical courses of action to address healthcare delivery gaps. This observation confirms the possibility of interactive effects between the two interventions.

## 2.6 Preliminary Pre-Analysis Plan for At-Scale Study

Based on the results of the pilot, this section presents a preliminary, condensed pre-analysis plan for a follow-on study with sufficient statistical power to rigorously evaluate the interventions. In general, the at-scale study will employ the same research design that has been described above for the pilot. To allow sufficient time to observe measurable effects on health services, the expanded study's intervention implementation period will last for approximately two years in between the baseline and endline data collection.

### 2.6.1 Research Questions

Along with quantitatively evaluating the overall impact of the programs, the at-scale study will enable several additional lines of inquiry. Specifically, the study will be designed to answer the following research questions:

1. *What are the causal effects of the meetings intervention, the training intervention, and the two interventions implemented together?*

---

<sup>15</sup>We do not, however, envision offering any monetary inducements, as we do not believe such an approach would be aligned with the program's focus on accountability and on the use of existing public resources more effectively.



This question will answer the most fundamental objective of the study and will elucidate the impact of the programs on politician behavior, health facility operations, health outcomes, and citizen satisfaction. A greater effect from the jointly implemented program may suggest complementarity between the two interventions.

2. *What are the mechanisms of the interventions? What underlying constraints do the programs alleviate?*

Based on the hypothesized theoretical framework and intervention mechanisms, this question will provide insight on the specific governance and institutional dynamics at play in the Ugandan health sector. The answers to this question may also point to additional promising points of intervention.

3. *Do the effects of the interventions depend on the competitiveness or functioning of the surrounding political environment in a given locality?*

This question will use the study's stratified design — based on recent margin of victory as a proxy for political competitiveness — to shed light on the influence of underlying political factors in producing a program impact.

4. *Do the interventions produce any negative externalities for localities neighboring the treated localities?*

If, in response to the interventions, local politicians begin lobbying more effectively on behalf of their communities to higher-level governing bodies (e.g., the district government), it is possible that resources allocated to the intervention localities will trade off with resources allocated to the non-intervention localities within the same legal jurisdiction. By carrying out data collection on resource allocations made to both treated and untreated localities within the same overall jurisdiction, the study will help to answer this question.

5. *Within the meetings arm, do changes in politician behavior depend on the number of persons who attend the meetings?*

Stemming from the observations in Section 2.4.2.1 that collective action plays a key role in political processes, this question will provide direct evidence on the influence

of collective action on politician behavior. Specifically, I will further randomize the localities within the meetings intervention group across different pre-meeting mobilization activities, based on approaches such as the encouragement messages studied in Chapter 1. The random assignment of mobilization procedures will create exogenous variation in the size of the meetings, enabling a credible evaluation of this question.

## 2.6.2 Statistical Framework

To answer the study’s primary question about the causal effects of the interventions, I will estimate the following regression equation:

$$Y_{ij} = \alpha + \beta_m M_j + \beta_s S_j + \beta_{ms} MS_j + u_{ij}, \quad (2.1)$$

where  $Y_{ij}$  denotes the outcome variable of interest for household  $i$  in locality  $j$ ,  $M_j$  is an indicator for the meetings only treatment,  $S_j$  is an indicator for the skills training only treatment, and  $MS_j$  is an indicator for the joint treatment. The coefficients of interest —  $\beta_m$ ,  $\beta_s$ , and  $\beta_{ms}$  — represent the intention-to-treat effect of the respective intervention. To support the internal validity of the experiment, I will test for pre-program balance and differential attrition between the control and treatment groups. All statistical analysis will use heteroskedasticity-robust standard errors adjusted for clustering at the locality level.<sup>16</sup>

## 2.6.3 Power Calculations

The data from the pilot enable power calculations to determine the necessary scale of the expanded cluster randomized controlled trial. Given the wide scope of data collected for the pilot, the data for many of the key outcomes that will be used in the at-scale study have not yet been fully processed. As more outcome data become available for analysis, I will carry out further power calculations, which will be presented in a complete version of the pre-analysis plan. At present, I perform power calculations for the following two citizen satisfaction outcomes from the household survey: satisfaction with health and satisfaction

---

<sup>16</sup>In a complete version of the pre-analysis plan, I will present fully specified hypothesis tests as well as fully specified regressions to evaluate the additional research questions outlined above.

with community, both of which are measured on a 1-5 Likert scale.

For each outcome of interest, the trial will test the null hypothesis  $H_0 : \mu_T = \mu_C$ , where  $\mu_T$  and  $\mu_C$  represent the means of the outcome for a given treatment group and the control group, respectively. Let  $\text{var}(\mu_T) = \sigma_T^2$  and  $\text{var}(\mu_C) = \sigma_C^2$ , and let the minimum detectable effect  $d = \mu_T - \mu_C$ . Based on this setup, I specify all of the power calculations for two-sided  $t$  tests between two populations, with significance level  $\alpha = 0.05$  and power  $1 - \beta = 0.8$  (where  $\beta = 0.2$  is the type II error probability). For a cluster randomized controlled trial with clusters of equal size, the required sample size per arm is given by the following equation (Hayes and Bennett 1999; Hemming et al. 2011; Hemming and Marsh 2013):

$$n = (\sigma_T^2 + \sigma_C^2) \left\{ \frac{(z_{\alpha/2} + z_{\beta})^2}{d^2} \right\} \left\{ 1 + (m - 1)\rho \right\}, \quad (2.2)$$

where the  $z$  values are determined using the inverse cumulative standard normal distribution at the specified probabilities,  $m$  is the fixed cluster size, and  $\rho$  is the intracluster correlation (ICC). In the foregoing equation, the quantity  $1 + (m - 1)\rho$  is known as the design effect, while the quantity  $(\sigma_T^2 + \sigma_C^2) \left\{ \frac{(z_{\alpha/2} + z_{\beta})^2}{d^2} \right\}$  gives the sample size that would be required under individual randomization (Donner, Birkett, and Buck 1981; Donner 1998). Accompanying equation (2.2), the required number of clusters per arm is given by

$$k = \left\lceil \frac{n}{m} \right\rceil + 1, \quad (2.3)$$

where the addition of one cluster is made to permit the use of the  $t$  distribution (Hayes and Bennett 1999).

For the power calculations, I set the mean of each variable to  $\mu_C$  and then choose  $d$ , which implies  $\mu_T$ . In addition, I assume that  $\sigma_T^2 = \sigma_C^2$ . After obtaining the minimum sample size and minimum number of clusters using equations (2.2) and (2.3), respectively, I make two adjustments. First, I round up the sample size so that it is a multiple of the cluster size. Second, I increase both the cluster size and the total sample size by ten percent to account for possible attrition between baseline and endline data collection.

In regard to the outcomes in question, the satisfaction with health variable has a mean of 3.546, a standard deviation of 1.367, and an ICC of 0.019, while the satisfaction with

community variable has a mean of 3.932, a standard deviation of 1.331, and an ICC of 0.000. For each of these outcomes, I consider a minimum detectable effect corresponding to a ten percent increase over the mean. Setting  $m = 30$  and applying all of the values to equation (2.2), I find that satisfaction with health is the limiting variable, requiring a larger sample size for the minimum detectable effect. Based on equations (2.2) and (2.3) alone, the study requires a sample size of 360 households across 13 clusters for each arm. After incorporating the additional aforementioned adjustments in the computation, I arrive at the final result of 429 households across 13 clusters for each experimental arm. Given the study’s four arms, these figures imply a total sample size of 1,716 households across 52 clusters. This projected sample size is relatively small. In reality, power calculations for the additional outcome variables of interest that have not yet been fully processed will likely yield a substantially larger required sample size.

## 2.7 Discussion

Many studies in development economics focus on the introduction of “novel” interventions implemented in clean, idealized settings; but the generalizability of these studies in terms of translating such interventions into national programs depends on background political economy factors (Acemoglu 2010; Deaton 2010). With potentially broad implications, this study instead directly targets political economy inefficiencies and structural governance factors that likely have pervasive incentive effects across multiple service sectors. Rather than introducing a new, often externally supported technology, the approach here focuses on using existing resources more efficiently and may be an especially promising strategy for sustained development.

In this fashion, the primary aim of this study has been to test the feasibility of two programs — community accountability meetings and politician training — along with corresponding data collection activities. Toward this end, the results demonstrate significant potential for both interventions to improve health service provision. For each of the programs, I record substantial ‘first-stage’ effects, with strong, representative attendance and participation achieved by the meetings and important learning outcomes registered by the

training. In addition, despite the complexities of many of the research instruments, all of the data collection activities proceeded smoothly. Beyond these outcomes, our team has also used the pilot to develop a rich set of implementation protocols and tools, creating a strong foundation for future work.

Building on these accomplishments and using the data from the pilot, I present an initial pre-analysis plan for a follow-on study conducted at-scale with sufficient statistical power. The pre-analysis plan outlines the main research questions of the scaled study, establishes the overall statistical framework for the measurement of program effects, and provides power calculations for the required sample size. The plan, which will be developed in greater depth as time goes on, offers a concrete roadmap for a high-quality quantitative evaluation.

Altogether, this study investigates supply-side barriers to quality healthcare delivery, contributing to the broader development economics literature on health interventions. Within this literature, many studies focus on demand-side barriers to health, addressing “puzzles” related to why poor households do not invest in health interventions that seem to produce significant private returns ([Dupas 2011b](#)). While demand-side barriers represent one part of the equation, supply-side factors are likely another essential side of the story. Indeed, it is possible that poor health service quality itself contributes to low demand for health services. By examining both politician capacity-building and citizen political participation in the context of the health sector, this study sheds light on the nature of governing institutions and important areas for reform. In addition, the meetings intervention in particular — by giving citizens a platform to demand improved health services, which in turn may lead to improved healthcare provision — connects both ends of the supply and demand equation.

At a macro level, multiple studies in economics highlight the role of institutions in development ([Acemoglu and Robinson 2001](#); [Hall and Jones 1999](#); [Acemoglu, Johnson, and Robinson 2002](#)). However, because of the aggregated nature of data, along with data quality concerns, reverse causality and omitted variable bias often complicate inference about the causal effect of institutions on development. The research design of both this pilot and the planned at-scale study, as randomized controlled trials using micro-level data, serves to provide more credible evidence on the impact of institutions on development.

The essence of the two programs in this study — with their focus on institutional quality

— is to remedy the current state of health and governance affairs by better operationalizing the law. How can we close the gap between “what is” and “what should be”? In establishing the framework for a larger, subsequent trial that rigorously evaluates the interventions and their associated mechanisms of action, this study contributes toward this key priority.

## References

- Abuya, Timothy, Charlotte E Warren, Nora Miller, Rebecca Njuki, Charity Ndwiga, Alice Maranga, Faith Mbehero, Anne Njeru, and Ben Bellows. 2015. “Exploring the Prevalence of Disrespect and Abuse during Childbirth in Kenya.” *PloS One* 10 (4): e0123606.
- Acemoglu, Daron. 2010. “Theory, General Equilibrium, and Political Economy in Development Economics.” *Journal of Economic Perspectives* 24 (3): 17–32.
- Acemoglu, Daron, Simon Johnson, and James A Robinson. 2002. “Reversal of Fortune: Geography and Institutions in the Making of the Modern World Income Distribution.” *Quarterly Journal of Economics* 117 (4): 1231–1294.
- Acemoglu, Daron, and James A Robinson. 2001. “The Colonial Origins of Comparative Development: An Empirical Investigation.” *American Economic Review* 91 (5): 1369–1401.
- Awortwi, Nicholas. 2010. “The Past, Present, and Future of Decentralisation in Africa: A Comparative Case Study of Local Government Development Trajectories of Ghana and Uganda.” *International Journal of Public Administration* 33 (12-13): 620–634.
- Bardhan, Pranab, and Dilip Mookherjee. 2006. “Decentralisation and Accountability in Infrastructure Delivery in Developing countries.” *Economic Journal* 116 (508): 101–127.
- Bidwell, Kelly, Katherine Casey, and Rachel Glennerster. 2020. “Debates: Voting and Expenditure Responses to Political Communication.” *Journal of Political Economy* 128 (8): 2880–2924.
- Bold, Tessa, and Jakob Svensson. 2013. “Policies and Institutions for Effective Service Delivery: The Need of a Microeconomic and Micropolitical Approach.” *Journal of African Economies* 22 (Supplement 2): ii16–ii38.
- Chaudhury, Nazmul, Jeffrey Hammer, Michael Kremer, Karthik Muralidharan, and F Halsey Rogers. 2006. “Missing in Action: Teacher and Health Worker Absence in Developing Countries.” *Journal of Economic Perspectives* 20 (1): 91–116.
- Chopra, Mickey, Elizabeth Mason, John Borrazzo, Harry Campbell, Igor Rudan, Li Liu, Robert E Black, and Zulfiqar A Bhutta. 2013. “Ending of Preventable Deaths from Pneumonia and Diarrhoea: An Achievable Goal.” *Lancet* 381 (9876): 1499–1506.
- Deaton, Angus. 2010. “Instruments, Randomization, and Learning about Development.” *Journal of Economic Literature* 48 (2): 424–455.
- Devas, Nick, and Ursula Grant. 2003. “Local Government Decision-Making—Citizen Participation and Local accountability: Some Evidence from Kenya and Uganda.” *Public Administration and Development* 23 (4): 307–316.
- Dizon-Ross, Rebecca, Pascaline Dupas, and Jonathan Robinson. 2017. “Governance and the Effectiveness of Public Health Subsidies: Evidence from Ghana, Kenya and Uganda.” *Journal of Public Economics* 156:150–169.
- Donner, Allan. 1998. “Some Aspects of the Design and Analysis of Cluster Randomization Trials.” *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 47 (1): 95–113.

- Donner, Allan, Nicholas Birkett, and Carol Buck. 1981. "Randomization by Cluster: Sample Size Requirements and Analysis." *American Journal of Epidemiology* 114 (6): 906–914.
- Dupas, Pascaline. 2011b. "Health Behavior in Developing Countries." *Annual Review of Economics* 3 (1): 425–449.
- Filmer, Deon, and Lant H Pritchett. 2001. "Estimating Wealth Effects without Expenditure Data—or Tears: An Application to Educational Enrollments in States of India." *Demography* 38 (1): 115–132.
- Fujiwara, Thomas, and Leonard Wantchekon. 2013. "Can Informed Public Deliberation Overcome Clientelism? Experimental Evidence from Benin." *American Economic Journal: Applied Economics* 5 (4): 241–55.
- Gottlieb, Jessica. 2016. "Greater Expectations: A Field Experiment to Improve Accountability in Mali." *American Journal of Political Science* 60 (1): 143–157.
- Grossman, Guy, and Kristin Michelitch. 2018. "Information Dissemination, Competitive Pressure, and Politician Performance between Elections: A Field Experiment in Uganda." *American Political Science Review* 112 (2): 280–301.
- Hall, Robert E, and Charles I Jones. 1999. "Why Do Some Countries Produce So Much More Output per Worker than Others?" *Quarterly Journal of Economics* 114 (1): 83–116.
- Hayes, RJ, and S Bennett. 1999. "Simple Sample Size Calculation for Cluster-Randomized Trials." *International journal of epidemiology* 28 (2): 319–326.
- Hemming, Karla, Alan J Girling, Alice J Sitch, Jennifer Marsh, and Richard J Lilford. 2011. "Sample Size Calculations for Cluster Randomised Controlled Trials with a Fixed Number of Clusters." *BMC Medical Research Methodology* 11 (1): 1–11.
- Hemming, Karla, and Jen Marsh. 2013. "A Menu-Driven Facility for Sample-Size Calculations in Cluster Randomized Controlled Trials." *The Stata Journal* 13 (1): 114–135.
- Hölmstrom, Bengt. 1979. "Moral Hazard and Observability." *Bell Journal of Economics*: 74–91.
- Kakumba, Umar. 2010. "Local Government Citizen Participation and Rural Development: Reflections on Uganda's Decentralization System." *International Review of Administrative Sciences* 76 (1): 171–186.
- Kiyaga-Nsubuga, John, and Yasin Olum. 2009. "Local Governance and Local Democracy in Uganda." *Commonwealth Journal of Local Governance*, no. 2: 26–43.
- Kyaddondo, David, Kidza Mugerwa, Josaphat Byamugisha, Olufemi T Oladapo, and Meghan A Bohren. 2017. "Expectations and Needs of Ugandan Women for Improved Quality of Childbirth Care in Health Facilities: A Qualitative Study." *International Journal of Gynecology & Obstetrics* 139 (S1): 38–46.
- Larreguy, Horacio, Benjamin Marx, Otis Reid, and Christopher Blattman. 2017. *A Market Equilibrium Approach to Reduce the Incidence of Vote-Buying: Evidence from Uganda*. Working Paper.
- Leonard, Kenneth L, and Melkiory C Masatu. 2010. "Using the Hawthorne Effect to Examine the Gap between a Doctor's Best Possible Practice and Actual Performance." *Journal of Development Economics* 93 (2): 226–234.



- Lewis, Maureen, and Gunilla Pettersson. 2009. *Governance in Health Care Delivery: Raising Performance*. Policy Research Working Paper 5074. Washington, DC: World Bank.
- Liu, Li, Shefali Oza, Daniel Hogan, Jamie Perin, Igor Rudan, Joy E Lawn, Simon Cousens, Colin Mathers, and Robert E Black. 2015. “Global, Regional, and National Causes of Child Mortality in 2000–13, with Projections to Inform Post-2015 Priorities: An Updated Systematic Analysis.” *Lancet* 385 (9966): 430–440.
- McPake, Barbara, Delius Asiimwe, Francis Mwesigye, Mathias Ofumbi, Lisbeth Ortenblad, Pieter Streefland, and Asaph Turinde. 1999. “Informal Economic Activities of Public Health Workers in Uganda: Implications for Quality and Accessibility of Care.” *Social Science & Medicine* 49 (7): 849–865.
- Olken, Benjamin A. 2007. “Monitoring Corruption: Evidence from a Field Experiment in Indonesia.” *Journal of Political Economy* 115 (2): 200–249.
- Paler, Laura. 2013. “Keeping the Public Purse: An Experiment in Windfalls, Taxes, and the Incentives to Restrain Government.” *American Political Science Review* 107 (4): 706–725.
- Pande, Rohini. 2011. “Can Informed Voters Enforce Better Governance? Experiments in Low-Income Democracies.” *Annual Review of Economics* 3 (1): 215–237.
- Pritchett, Lant, and Michael Woolcock. 2004. “Solutions when the Solution is the Problem: Arraying the Disarray in Development.” *World Development* 32 (2): 191–212.
- Raffler, Pia. 2020. *Does Political Oversight of the Bureaucracy Increase Accountability? Field Experimental Evidence from a Dominant Party Regime*. Working paper.
- Ravanilla, Nico. 2016. “Attracting Good People into Public Service: Evidence from a Field Experiment in the Philippines.” Unpublished.
- Republic of Uganda. 2000. “Laws of Uganda.” Chap. 243: The Local Governments Act. Kampala, Uganda: Republic of Uganda.
- Ross, Stephen A. 1973. “The Economic Theory of Agency: The Principal’s Problem.” *American Economic Review* 63 (2): 134–139.
- Sharma, Gaurav, Matthews Mathai, Kim E Dickson, Andrew Weeks, G Justus Hofmeyr, Tina Lavender, Louise Tina Day, Jiji Elizabeth Mathews, Sue Fawcus, Aline Simen-Kapeu, et al. 2015. “Quality Care during Labour and Birth: A Multi-Country Analysis of Health System Bottlenecks and Potential Solutions.” *BMC Pregnancy and Childbirth* 15 (2): S2.
- Tripp, Aili Mari. 2010. *Museveni’s Uganda: Paradoxes of Power in a Hybrid Regime*. Vol. 6. Lynne Rienner Publishers Boulder, CO.
- UBOS and ICF. 2012. *Uganda Demographic and Health Survey 2011*. Country Report. Kampala, Uganda and Calverton, Maryland: Uganda Bureau of Statistics and ICF International Inc.
- . 2018. *Uganda Demographic and Health Survey 2016*. Country Report. Kampala, Uganda and Rockville, Maryland: Uganda Bureau of Statistics and ICF International Inc.

- UN IGME. 2020. *Levels and Trends in Child Mortality: Report 2020*. Technical Report. New York, NY: UN Inter-agency Group for Child Mortality Estimation.
- World Bank. 2016. *Health Service Delivery in Uganda*. Technical Report AUS5359. Washington, DC: World Bank.
- . 2003. *World Development Report 2004: Making Services Work for Poor People*. Washington, DC: World Bank.
- . 2017. *World Development Report 2017: Governance and the Law*. Washington, DC: World Bank. doi:[10.1596/978-1-4648-0950-7](https://doi.org/10.1596/978-1-4648-0950-7).

## CHAPTER 3

### Demand-Side Factors in Maternal Health Outcomes: Evidence from a Community Health Worker Program in Uganda

#### 3.1 Introduction

##### 3.1.1 Background

In recent years, community health workers (CHWs) have become a core feature of many low-resource healthcare systems and have been lauded for their ability to target vulnerable populations, better connect patients to local health facilities, and provide community-level care. Along these lines, CHW programs may have a role to play in a range of primary health-care functions, including maternity care. This study evaluates one such CHW intervention focused on improving maternal health in southwestern Uganda. As in many low-income countries, maternal mortality in Uganda persists at an alarming level, with an estimated 375 per 100,000 women dying in childbirth each year ([WHO et al. 2019](#)). Closely related to maternal health outcomes, the perinatal mortality rate stands at 38 per 1,000 pregnancies of seven or more months' duration ([UBOS and ICF 2018](#)). These figures, along with a substantial burden of maternal and infant morbidity, are thought to be driven by factors such as suboptimal rates of health facility deliveries and inadequate levels of antenatal and postnatal care attendance. For example, according to data from shortly before the start of this study, only 40.3 percent of women in southwestern Uganda gave birth in a health facility, with the remainder largely giving birth at home without any skilled attendance ([UBOS and ICF 2012](#)). Although the rate of skilled deliveries has since risen, more recent data show that about 50 percent of Ugandan women do not have their first antenatal care visit until about the fifth month of pregnancy, while about one-third of women complete only 2-3 such

visits ([UBOS and ICF 2018](#)).

At an individual level, the benefits of maternity care appear significant enough to far outweigh the costs. Moreover, at a societal level, there is widespread agreement on the developmental importance of maternal health, which amounts to the earliest possible form of childhood health intervention. As a common refrain in the development economics literature on health goes, why does demand for health services by the poor remain low, given that investing in these services seems to have high private welfare returns? If the benefits of maternity care outweigh the costs, low levels of maternal healthcare in low- and middle-income countries suggest departures from the predictions of neoclassical economic behavior.

In this study, I examine these possible departures by evaluating a CHW program in rural Uganda to shed light on the incentives and constraints that influence health choices during pregnancy. Carried out by the organization Progressive Health Partnership (PHP), the program trained members of the target community to work as CHWs, visiting women during and after pregnancy to provide basic health checks and advice. Employing a difference-in-differences estimation technique, I find relatively little evidence of an overall program effect on health behaviors, including antenatal care attendance and delivery under skilled supervision. Analysis of heterogeneity by gestational age at first antenatal visit — which should modulate exposure to the intervention — provides suggestive evidence that treatment effects vary predictably with gestational age. Analysis of heterogeneity based on CHW gender and location points to a few possible treatment effects as well. In contrast to the weak findings on program effects, I find large improvements in healthcare-seeking behavior in the study sample between the pre- and post-intervention periods. These changes may arise from concurrent health facility improvements affecting the entire study population, spillover effects from the CHWs, or background health trends. If the CHW intervention produced no effect, one explanation may be suboptimal performance by CHWs, thus highlighting the importance of studying and instituting appropriate management measures and incentives for such programs.

### 3.1.2 Literature Review

Motivating the CHW program, imperfect information and intra-household bargaining may be important in explaining observed departures from neoclassical economic predictions in a variety of household decisions and may therefore provide promising points of intervention. In particular, several studies have demonstrated changes in household decision-making in response to both information on health risks and information on specific prevention techniques.<sup>1</sup> However, the effectiveness of information provision may depend on its level of credibility and the characteristics of the intended recipient (Dupas 2011b). Information also may not be effective if the targeted recipients face preceding constraints in their lives. This study contributes to the literature by examining the effects of information provision for pregnant women. Specifically, the study provides an opportunity to assess the credibility of CHWs in providing this health information.

In regard to intra-household decision-making, Ashraf, Field, and Lee (2014) find important effects of gender-specific preferences and information on adoption of contraception. Several studies also specifically address women’s bargaining power in relation to maternal health behaviors.<sup>2</sup> However, few studies causally identify effects that may operate via changes in intra-household bargaining (Doss 2013). The present study helps to address this gap by providing a source of causal identification. As a program evaluation, it also differs from much of the previous literature by focusing on a possible strategy to directly address low bargaining power on the part of women. In addition, in contrast to many suggested strategies to increase women’s power within the household, such as increasing women’s income or asset ownership, the intervention here addresses norms as constraints on women’s bargaining power — that is, the common practices of couples in relation to maternity care.

From a policy standpoint, utilizing CHWs to enhance maternal healthcare holds great promise for improving outcomes. Recent years have seen an increasing focus on formally integrating CHWs into low-resource healthcare systems, with countries such as Brazil, Pak-

---

<sup>1</sup>For example, see Cohen, Dupas, and Schaner (2015); Dupas (2011); Jalan and Somanathan (2008); Madajewicz et al. (2007); Cairncross et al. (2005) and Rhee et al. (2005).

<sup>2</sup>For example, see Peters (2011); Maitra (2004); Beegle, Frankenberg, and Thomas (2001); Allendorf (2007); Hou and Ma (2013); Danforth et al. (2009) and Becker, Fonseca-Becker, and Schenck-Yglesias (2006).

istan, Ethiopia, and India making CHWs central to their efforts to meet the health needs of their populations. For the health priorities of sub-Saharan Africa in particular, multiple international bodies have called for a major expansion of CHW programs ([Singh and Sachs 2013](#)), and the World Health Organization recommends the active involvement of CHWs in maternity care ([WHO 2012](#)). However, despite these calls, researchers have produced varying degrees of evidence on the effectiveness of CHWs in different settings.

Only a few studies, all situated in the public health literature, have employed rigorous research methods to credibly measure the causal effect of CHW home visits on maternal health outcomes in sub-Saharan Africa. Using a randomized controlled trial, [Kirkwood et al. \(2013\)](#) show that a CHW home visit program in Ghana leads to improvements in a range of newborn care behaviors but does not affect the important indicators of antenatal care visit completion and delivery with skilled attendance. [Waiswa et al. \(2015\)](#) produce analogous results with a similar study design in Uganda. In contrast, [Geldsetzer et al. \(2019\)](#), also using an experimental approach, find an increase in skilled deliveries but no change in the completion of four antenatal visits under a CHW intervention in Tanzania. A couple additional trials have provided evidence for a positive effect of CHWs on maternal health outcomes, but they have suffered from significant methodological weaknesses.<sup>3</sup> Differing from the foregoing studies, the present study provides a quasi-experimental context in which to rigorously evaluate the impact of CHWs on maternal health. In addition, the study is specifically designed around the dimensions of economic behavior that may influence the success of the CHW intervention.

---

<sup>3</sup>[Akinyi, Nzanzu, and Kaseje \(2015\)](#) report an increase in antenatal care attendance and health facility deliveries, but their control group originates from an entirely different district. Without data to assess the balance between the treatment and control groups, we are unable to determine whether the two groups systematically differ from one another. In addition, their data collection uses reports from health facilities and CHWs, which are likely to be biased. [Adam et al. \(2014\)](#) also report positive effects on facility deliveries among women exposed to CHWs. However, the control group in their study was composed of women who themselves chose to attend a teaching by a CHW, suggesting a high likelihood of selection bias. Indeed, the post-intervention demographic data that the authors present appear to confirm this possibility, with significant imbalances for several characteristics, at least some of which presumably existed at baseline.

### 3.1.3 Theoretical Framework

Drawing on the above discussion, I envision two primary pathways through which the CHW program may operate: information provision and bargaining power. First, because of deficiencies in education, as well as an overall lack of programs to raise public awareness, women and their husbands may experience imperfect information regarding the private returns to investment in maternal health services. By offering personalized health education and advice to women and their husbands in their homes, CHWs may incentivize greater take-up of maternal health services.

On one hand, CHWs, as fellow community members, may occupy a unique social position as trusted neighbors, in contrast to the formal healthcare system which tends to be an object of mistrust in Uganda. Additionally, unlike local government health workers, CHWs operate at the village and household levels, making health advice easier to access. Given these considerations, CHWs may serve as a more optimal mechanism for the delivery of information. On the other hand, women and families may view CHWs less credibly, perhaps due to their lack of formal health training or to observations that the CHWs themselves do not follow their health advice in their own lives. In this case, CHWs may have little impact on maternal health behaviors.

Second, low intra-household bargaining power of women may inhibit their ability to make optimal health choices. Women may experience low bargaining power because of imperfect information on the part of husbands in regard to the benefits their support can achieve or because of different male preferences in regard to maternal health choices. In turn, low bargaining power may lead to financial constraints that hamper a woman's ability to overcome barriers to care, such as transportation to a health facility when in labor. In addition, because women often bear the responsibility for most household chores — even during pregnancy — the opportunity cost of visiting a health center for antenatal or postnatal care may exceed the perceived benefits.

Following from these observations, intra-household bargaining effects may exhibit heterogeneity based on household characteristics. For instance, for households that reside closer to the local health center, the opportunity cost of visiting the health facility is lower, and

therefore women may not be as constrained by limited bargaining power. Similarly, for wealthier households, bargaining power may not be a major constraint for women, since a higher aggregate level of resources is available to distribute among household members. Aside from heterogeneity in potential effects due to intra-household bargaining dynamics, it is also possible that women’s level of bargaining power itself varies based on demographic characteristics. If, for example, a household has experienced the death of one or more children, women may have higher bargaining power in relation to maternal health, insofar as their husbands recognize the importance of maternal decision-making to help ensure the birth of healthy children.

Importantly, intra-household bargaining dynamics may modulate the effects of information provision on maternal health behaviors. If the CHW intervention were targeted only to women, the effect of information provision may remain constrained by unaltered bargaining processes. However, by speaking with women and their husbands about the role that each party plays over the course of pregnancy and encouraging men to become more supportive of their wives’ care, CHWs may serve to increase the bargaining power of women. Specifically, this mechanism of action runs counter to gender norms, which lead few men to be involved in the process of seeking maternity care. Thus, any impact of the intervention that operates through a bargaining pathway should occur via the intervention’s effects on a couple’s conformity with norms.

## **3.2 Program Background and Data Source**

### **3.2.1 Study Location and Population**

This study uses data from an evaluation of the CHW intervention run by PHP, a non-profit organization that collaborates with the government in southwestern Uganda to improve healthcare delivery. PHP operates in two neighboring rural subcounties, Kashongi and Kitura, which together encompass thirteen parishes<sup>4</sup>, 106 villages, and approximately 41,000 people in Kiruhura District. The catchment area spans several square miles. Residents are

---

<sup>4</sup>Parishes are geographic divisions, each of which is further divided into several villages.



predominantly poor, and most earn a living from subsistence farming.

### **3.2.2 Program Description and Study Arms**

In 2012, PHP launched a new maternal and neonatal health initiative, the Omukazi Namagara Program, in its 106 partner villages. The program consisted of two components: (1) quality of care improvements at public health facilities (including more reliable stocks of medications and supplies, provision of new equipment, increased staffing, and improved patient education) and (2) home visits to women during and after pregnancy conducted by PHP-trained CHWs.

#### **3.2.2.1 Treatment and Control Groups**

PHP implemented the first component of the program at the Kashongi and Kitura Health Center III's, the two main government health centers that serve the subcounties. As a result, all members of the 106 villages had access to the improved facility-level services. However, the organization only implemented the second component of the program in 52 of the 106 villages, so only women in these villages received home visits from a CHW. I hereafter refer to these 52 villages as the treatment villages and the remaining 54 villages as the controls. It is the second component of the program — the CHW home visits — that is the main subject of this study. In addition to the two abovementioned program components, at the start of the program PHP conducted an information campaign throughout all 106 villages to raise awareness of various aspects of maternal and neonatal health.

#### **3.2.2.2 Community Health Worker Background and Selection**

PHP selected the CHWs for the program through the Ugandan government's Village Health Team. The Village Health Team generally consists of multiple individuals in each village whom the government has identified and recruited to serve as CHWs on a voluntary basis. In each subcounty, the government has designated one member of the Village Health Team as the subcounty coordinator. Similarly, in each parish within each subcounty, the government has designated parish coordinators. Subcounty coordinators dual-serve as parish

coordinators in their home parishes.

Over the duration of the study, however, the Village Health Team members were largely inactive, absent third-party activity by an organization such as PHP. For the Omukazi Namagara Program, PHP identified a subset of Village Health Team members with whom to work, since working with the full roster of Village Health Team members would be infeasible. Specifically, the organization selected two Village Health Team members for each of the thirteen parishes in Kashongi and Kitura, making for a total of twenty-six Village Health Team members on its team. In each parish, PHP selected one male and one female CHW so that the team would be evenly split by gender.

The organization's selection of CHWs was driven by several considerations. First, although it would only be working with a subset of the Village Health Team members, PHP sought to maintain the core structure that the government had intended for the Village Health Team. As a result, the organization chose both of the subcounty coordinators and all of the parish coordinators for its team. Second, beyond these thirteen CHWs, PHP had existing relationships from previous work in the communities with five Village Health Team members whom it added to the team. Third, the organization randomly selected the remaining eight CHWs.

### **3.2.2.3 Community Health Worker Village Assignments**

PHP assigned each of the CHWs to work in two villages: the home village of the CHW and the nearest neighboring village, leading to 52 treatment villages total. Although both the CHW and village selection processes meant that not all of the treatment villages were randomly selected, their distribution across the entire catchment area appears to be relatively dispersed. The selection of CHWs by both the government and PHP was driven more than anything by idiosyncratic circumstances related to the qualities of these individuals and their history of leadership in the two subcounties, rather than by any defining features of their home villages. Also, in spite of the organization's intention to maintain the structure of the Village Health Team, in practice the CHWs operated with relatively little hierarchy.

#### **3.2.2.4 Community Health Worker Training, Compensation, and Oversight**

Using materials that the organization designed, PHP trained the team of CHWs on a range of topics related to maternal health, such as the importance of receiving care in a health facility, antenatal and postnatal hygiene, sanitation, nutrition, malaria prevention, basic recognition of danger signs during and after pregnancy, referral of women to the local health centers, basic information on antenatal and postnatal medications and vitamins, and appropriate breastfeeding and immunization practices. The CHWs also received training on interpersonal skills and on social issues that may complicate maternity care. PHP provided specific training on involving husbands in maternity care and instructed the CHWs to talk with both a woman and her husband (if applicable) when making a home visit. All CHWs received tools to guide their home visits and recordkeeping. Following the training, PHP administered a competency exam to the CHWs.

According to the program guidelines, CHWs were supposed to make approximately three home visits during the antenatal period and two home visits during the postnatal period for each woman. PHP paid the CHWs 5,000 UGX for each home visit made. After the CHWs began the program activities, PHP monitored and oversaw their work mainly through periodic in-person assessments and through quarterly meetings held with the entire team.

#### **3.2.3 Pre-Intervention Survey**

At the beginning of the program in July 2012 and lasting through July 2013, PHP conducted a baseline survey of all pregnant women who were attending their first antenatal care visits at each of the Health Center III's. The sample included 1,131 women in total. Although this sampling technique implies that women who never access antenatal care would be excluded from the sample, such population-level selection bias is likely to be minor, since data from prior to the start of the study indicated that 95.4 percent of women in rural Uganda attend at least one antenatal care visit over the course of pregnancy ([UBOS and ICF 2012](#)). In the study area, those women who obtain antenatal care are likely to do so via one of the government health centers at which PHP carried out the pre-intervention survey.

The data collection team collaborated with the midwives at the Health Center III's in an

effort to integrate the interviews into the normal work flow of the health centers as seamlessly as possible. As each health center had designated days of the week for antenatal care, the data collection team also aimed to distribute itself optimally across the health centers so as to maximize the number of women interviewed. On many of the days during the baseline interview period, the data collection team surveyed all of the women present for antenatal care. However, on some days it was not feasible to interview all of the women, as the number of women who came for antenatal services may have been large relative to the number of interviewers. No woman who was asked to participate in the survey refused.

The survey lasted 1-1.5 hours and covered topics such as individual and household demographics; health practices and outcomes during previous pregnancies; care sought and plans for the current pregnancy; and attitudes, knowledge, and practices related to HIV/AIDS, malaria, and family planning. In the treatment villages, each woman interviewed for the survey only began receiving CHW home visits after her interview.

### **3.2.4 Post-Intervention Survey**

Between March 2014 and March 2015, PHP conducted a follow-up survey of the same sample of women, this time interviewing them at their homes. While organizing the logistics of the follow-up survey as efficiently as possible, the data collection team aimed to re-interview the women sequentially according to the time at which they were interviewed during the baseline survey. This procedure ensured that approximately the same duration of time elapsed between each woman's pre- and post-intervention interview. The content of the follow-up survey adhered to a similar structure as the baseline survey. However, despite considerable follow-up efforts, there was significant attrition of the sample, with only 725 women re-interviewed.

## 3.3 Empirical Approach

### 3.3.1 Outcomes of Interest

In evaluating the effect of the CHW home visits, I focus on the following outcome variables: (1) attendance of four or more antenatal care visits<sup>5</sup>; (2) delivery with skilled attendance<sup>6</sup>; (3) breastfeeding within thirty minutes of birth; and (4) improper umbilical cord care, which is defined as putting anything on the newborn’s umbilical cord beside clean soap and water. The first two outcomes capture key targets used in measuring a woman’s access to the formal healthcare system for maternity care, while the latter two outcomes give measures of practices known to have direct impacts on health.

### 3.3.2 Difference-in-Differences Specification

To uncover the program’s intention-to-treat (ITT) effects, I employ a difference-in-differences framework, given by the following specification:

$$Y_{ijt} = \beta_0 + \beta_1 T_j + \beta_2 POST_t + \beta_3 (T_j * POST_t) + u_{ijt}, \quad (3.1)$$

where  $Y_{ijt}$  denotes the level of the outcome variable of interest for woman  $i$  in village  $j$  at time period  $t$ ,  $T_j$  is an indicator variable that is equal to 1 if village  $j$  is a member of the treatment group,  $POST_t$  is an indicator variable that is equal to 1 for the time period after the intervention, and  $u_{ijt}$  is an error term. The coefficient of interest is  $\beta_3$ , which represents the ITT of the program.  $\beta_1$ , the coefficient on the treatment variable, represents the group fixed effect of being in the treatment arm, while  $\beta_2$ , the coefficient on the post-program variable, represents the time fixed effect of the period following the introduction of the program. For all of the analysis in this article, I use heteroskedasticity-robust standard errors that account for clustering at the village level.

---

<sup>5</sup>Although the Uganda Ministry of Health recently increased the recommended number of antenatal care visits from four to eight, many women in Uganda continue to have four or less visits.

<sup>6</sup>Skilled attendance almost certainly implies that a delivery occurs in a professional health facility. However, a home delivery with the attendance of a licensed health professional would also qualify as a skilled delivery. In the study sample, only one woman reported such a case.

In the baseline survey, all women were asked about their practices during their most recent previous pregnancy within the past five years (if applicable) apart from the current one. The reported practices for the most recent previous pregnancy serve as the baseline levels that I use for the outcome variables in the difference-in-differences specification. Because this procedure excludes primigravida (first-time pregnant) women, I employ additional specifications shown below to account for the effects on this sub-population (see Section 3.4.4).

The coefficient  $\beta_3$  represents the differential pre-post change in the outcome variable in the treatment group compared to the pre-post change in the outcome variable in the control group. For  $\beta_3$  to have a causal interpretation, the model contains the parallel trends assumption. Letting  $Y_{ijt}(T_j)$  denote an individual's potential outcome, the assumption requires that  $\mathbb{E}[Y_{ij1}(0) - Y_{ij0}(0) | T_j = 0] = \mathbb{E}[Y_{ij1}(0) - Y_{ij0}(0) | T_j = 1]$ . In words, the average pre-post change in potential outcomes for the control group must equal the average pre-post change in potential outcomes that would have occurred in the treatment group had the program not been implemented.

### 3.3.3 Attrition

Reports from the data collection team indicated that the high level of sample attrition arose because a large number of women in the pre-intervention survey came from outside of Kashongi and Kitura but misrepresented their residential location, thinking that they needed to be from Kashongi or Kitura to be eligible for PHP's services at the health centers (a belief that was untrue). Consequently, it was exceedingly difficult to relocate such women at follow-up. Migration accounted for the other primary source of attrition.

For difference-in-differences estimation with a balanced panel of respondents, attrition does not introduce any bias into the identification result. Even if attrition is systematically related to membership in the treatment or control group, it has no effect on the parallel trends assumption, which is the only condition needed for the internal validity of the estimation strategy. Naturally, however, the group of attrited respondents differs from the non-attrited portion of the sample along a number of dimensions, which may call into question the external validity of the results. Table 3.1 documents these differences, showing the mean levels of a range of baseline variables for the full sample, missing respondents, and non-

missing respondents. The last column of the table gives  $p$ -values for difference-in-means tests across the missing and non-missing groups. The bottom of the table displays the  $p$ -value from a joint test of differences, which is highly statistically significant. Among other findings, attrited respondents have lower gravidity, are more likely to have received antenatal services from another source, are more likely to use herbs, are less likely to have husbands who make their health decisions, are more likely to have had a recent skilled delivery, and are less likely to have breastfed within 30 minutes of birth at their most recent delivery. While the parallel trends assumption is unaffected by sample attrition, note that the last test in the table verifies that the attrition is not systematically related to membership in the treatment or control group. All further analysis in this article uses only the balanced panel of non-attrited respondents, which I refer to as the “full” sample.

### 3.3.4 Analysis of Pre-Program Trends

Although the parallel trends assumption itself is not testable (since the potential outcomes under control conditions are not observed for the treatment group), I examine pre-program trends in the outcome variables across the treatment and control groups to help justify the assumption. This analysis uses two primary approaches: (1) I perform placebo tests for each outcome variable in the pre-program period to verify that a false treatment would not have been associated with a statistically significant divergence in trends between the treatment and control groups; and (2) I estimate the control and treatment linear time trends for each outcome variable in the pre-program period and test for equality of trends across the study arms.

To construct the pre-program trends, I take advantage of the data on each woman’s most recent previous pregnancy within the past five years. Using either the approximate reported age of the child from this pregnancy or the approximate number of years elapsed since the pregnancy if the child did not survive<sup>7</sup>, I back out the approximate number of years prior to baseline when this pregnancy occurred. Thus, in each year  $t \in \{-5, -4, -3, -2, -1\}$ , I observe the outcomes for the subset of women in the sample who gave birth in that year.

---

<sup>7</sup>For cases in which the child did not survive, women were asked about any pregnancies that lasted at least six months from which the child was no longer alive at the time of the baseline interview.

TABLE 3.1: ANALYSIS OF SAMPLE ATTRITION

	(1)	(2)	(3)	(4)
	Overall	Missing	Non-Missing	<i>p</i> -value
Respondent Education Level	5.863	5.987	5.793	0.293
Household Head Education Level	6.320	6.468	6.240	0.329
Respondent is Household Head	0.030	0.037	0.026	0.308
Asset Index	-0.000	-0.045	0.025	0.584
Frequency of Listening to Radio	1.915	1.995	1.871	0.096
Household Owns Mosquito Bednet	0.750	0.686	0.786	0.000
Family Member Maternal/Neonatal Death	0.451	0.480	0.435	0.118
Number of Previous Pregnancies	2.752	2.015	3.165	0.000
Gestational Age (weeks)	18.988	19.938	18.452	0.006
Walked to Health Center	0.596	0.601	0.594	0.813
Accompanied by Husband to Health Center	0.134	0.155	0.123	0.131
Time to Health Center	67.720	66.997	68.126	0.706
Burden to Come to ANC	0.509	0.519	0.503	0.654
Received ANC from Other Source	0.222	0.303	0.176	0.000
Has Taken Herbs	0.539	0.594	0.509	0.004
Number of HIV Prevention Methods Known	1.681	1.712	1.664	0.365
Husband Makes Health Decisions	0.398	0.345	0.427	0.009
Four ANC Visits	0.564	0.569	0.562	0.872
Skilled Delivery	0.490	0.591	0.451	0.003
Husb. Decides Location	0.267	0.273	0.264	0.834
Breastfeed 30 Minutes	0.642	0.570	0.670	0.005
Improper Cord Care	0.646	0.674	0.636	0.372
Treatment	0.607	0.641	0.588	0.160
Joint ( $\chi^2$ ) test <i>p</i> -value				0.000

*Notes.* This table displays baseline covariate means for the overall sample and for the sub-samples that were missing and non-missing at endline, along with *p*-values for difference-in-means tests between these sub-samples. All tests use heteroskedasticity-robust standard errors adjusted for clustering at the village level.

Appendix C.2 contains graphs of the trends across these years for each outcome variable of interest. As it is not possible to construct pre-program trends for the full sample, the pre-program trends here constitute trends across repeated cross sections rather than a panel of subjects. In each year prior to the baseline survey, we observe different groups of women, which may differ from one another based on key characteristics. Despite this limitation, the pre-program trends that I construct capture important features in the evolution of the sample over time, and the results of the tests that follow lend confidence to the parallel trends assumption.

Using the difference-in-differences specification in equation (3.1), I perform falsification



tests between each consecutive pair of years  $t$  and  $t + 1$  (for  $t + 1 \leq -1$ ), imagining year  $t + 1$  as the post-program period. Because of a particularly small sample size in the period five years preceding the baseline, I omit the interval between five and four years prior to baseline from testing. Table 3.2 shows the results of these pre-program falsification tests, with each panel displaying the estimates of the difference-in-differences coefficient,  $\beta_3$ , in the corresponding regressions. A statistically significant estimate of  $\beta_3$  for any of the year pairs would indicate non-parallel trends. This coefficient estimate is not statistically significant at conventional levels in any of the tests.

Supplementing these findings, Appendix C.3 contains results from a pooled version of the falsification test along with tests of equality in linear time trends between the treatment and control groups. Both sets of results corroborate the findings shown in Table 3.2. Thus, I maintain each of the outcome variables of interest in the analysis that follows.

TABLE 3.2: PRE-PROGRAM FALSIFICATION TESTS

	-4 to -3 Years	-3 to -2 Years	-2 to -1 Years
<i>Panel A: Four ANC Visits</i>			
Post * Treatment	0.166 (0.171)	0.056 (0.120)	-0.036 (0.121)
<i>Panel B: Skilled Delivery</i>			
Post * Treatment	0.039 (0.167)	0.042 (0.123)	-0.152 (0.123)
<i>Panel C: Breastfeed 30 Minutes</i>			
Post * Treatment	-0.095 (0.176)	0.070 (0.118)	0.102 (0.112)
<i>Panel D: Improper Cord Care</i>			
Post * Treatment	0.018 (0.162)	-0.172 (0.124)	0.038 (0.125)

*Notes:* This table displays the results of pre-program falsification tests between successive years prior to the baseline survey. Heteroskedasticity-robust standard errors, adjusted for clustering at the village level, are shown in parentheses. \*Significant at 10% level, \*\*Significant at 5% level, \*\*\*Significant at 1% level.

### 3.3.5 Assessment of Concurrent Programs or Policy Changes

In addition to the null results from the pre-program trends analysis, I am unaware of any other major programs or policy changes related to maternal health that occurred in Kashongi and/or Kitura during the study period and that would have differentially affected the treat-

ment and control groups. Because PHP works closely with the government, the organization would have been informed of any significant changes in healthcare delivery activities. Outside of the government, the organization is familiar with two other non-governmental organizations that work in Kashongi and/or Kitura. One focuses mainly on child development and education, while the other operates a broad-based community center that includes some health, but no maternal health, activities. The programs of both organizations remained stable and relatively unchanged throughout the study period. Altogether, this landscape suggests that the treatment and control groups likely would have continued to evolve in parallel had PHP’s program not been implemented.

## 3.4 Results

### 3.4.1 Intention-to-Treat Effects

Using equation (3.1), I estimate the ITT effect — the average effect of being assigned to treatment — of the CHW home visits on each of the outcome variables of interest. Table 3.3 displays the results. The third row of the table, for the interaction term, gives the estimate of the coefficient of interest,  $\beta_3$ . The bottom two rows of the table show the mean counterfactual level of the dependent variable and the post-program mean of the dependent variable in the control group. The mean counterfactual level, equal to the estimate of  $\beta_0 + \beta_1 + \beta_2$ , gives the average level of the outcome variable that the treatment group would have experienced had the parallel time trend continued and the program not been introduced. The control group post-program mean is given by the estimate of  $\beta_0 + \beta_2$ .

None of the outcome variables show statistically significant treatment effects, although several of the estimates are worth examining. The estimated effect on completing four antenatal care visits is unexpectedly negative, with the treatment group experiencing a differential change of -5.6 percentage points relative to the control group. All of the other treatment effect estimates move in the expected direction. The treatment group experienced a 4.0 percentage point differential increase in the likelihood of giving birth with skilled attendance. Also, improper cord care exhibited a 9.0 percentage point differential pre-post

TABLE 3.3: INTENTION-TO-TREAT EFFECT ESTIMATION

	(1)	(2)	(3)	(4)
	Four ANC Visits	Skilled Delivery	Improper Cord Care	Breastfeed 30 Minutes
Treatment	0.117** (0.047)	0.012 (0.050)	0.086* (0.050)	-0.005 (0.043)
Post	0.187*** (0.038)	0.085** (0.035)	0.060 (0.047)	0.021 (0.048)
Post * Treatment	-0.055 (0.056)	0.040 (0.048)	-0.090 (0.061)	0.017 (0.060)
Constant	0.493*** (0.037)	0.443*** (0.039)	0.583*** (0.039)	0.686*** (0.032)
Mean Counterfactual	0.796	0.540	0.729	0.701
Mean Control Post	0.679	0.528	0.643	0.707

Notes: This table displays the results of the main intention-to-treat estimation, with the coefficient of interest corresponding to the  $Post * Treatment$  variable. Heteroskedasticity-robust standard errors, adjusted for clustering at the village level, are shown in parentheses. \*Significant at 10% level, \*\*Significant at 5% level, \*\*\*Significant at 1% level.

decrease in the treatment group relative to the control, an effect that approaches statistical significance and is sizable compared to the mean counterfactual level of 72.9 percent. The estimate of  $\beta_3$  for breastfeeding within 30 minutes of birth is minimal.

For additional context, note that the interpretation of the parameter of interest as an ITT in these estimations arises not because of limited intervention adoption by pregnant women (e.g., refusal to allow a CHW to conduct home visits) but because of limited intervention delivery by CHWs. Although PHP aimed for the CHWs to visit all pregnant women in their assigned villages, only 59.2 percent of women in the treatment group reported receiving a home visit.

### 3.4.2 Heterogeneity by CHW Characteristics

While the treatment group does not appear to experience an effect due to the intervention, it is possible that the treatment effect depends on CHW characteristics. Location and gender are two such features that may be important. As described above, each CHW was assigned to work in his or her home village and the nearest neighboring village. However, the workload of two villages may have been overly demanding. In addition, CHWs reported difficulties to PHP about working in neighboring villages due to lesser familiarity with all of the households. Thus, it is possible that treatment intensity is lower in neighboring villages. According to the data, 62.9 percent of women in CHW home villages receive at least one home visit, compared to 55.0 percent of women in CHW neighboring villages — a modest difference that is not statistically significant. Among those who receive at least one home visit, women in CHW home villages report an average of 2.26 visits during pregnancy, compared to an average of 2.0 visits in CHW neighboring villages ( $p = 0.106$ ).

Treatment quality or intensity may also differ by CHW gender. While the precise role of gender is not clear and may vary across outcomes, there are a number of possibilities. For example, female CHWs may be more effective at communicating with pregnant women. Alternatively, male CHWs may be more effective at fostering the involvement of male partners. Gender may also be associated with the CHW's own health beliefs or practices, which he or she may convey to the program participant.

To explore the existence of such differential treatment effects, I estimate the following

equation, decomposing the treatment group into two sub-groups based on either CHW location or gender:

$$Y_{ijt} = \beta_0 + \beta_1 T1_j + \beta_2 T2_j + \beta_3 POST_t + \beta_4 (T1_j * POST_t) + \beta_5 (T2_j * POST_t) + u_{ijt} \quad (3.2)$$

Because this estimation again requires the parallel trends assumption, I conduct pre-treatment trends tests for each pair of decomposed treatment groups. Displaying the falsification test results, Tables 3.4 and 3.5 do not indicate any discrepant trends, with the exception of the skilled delivery outcome for the female CHW treatment group in the period immediately preceding the baseline survey. This evidence for a potential violation of the parallel trends assumption, however, does not substantively impact any conclusions because no program effect occurs for this same outcome and group. Appendix C.3 also contains results for the pooled falsification tests and equality in time trends tests, which do not demonstrate divergent trends between any of the treatment groups and the control.

#### 3.4.2.1 CHW Location

Table 3.6 shows the results for the estimation differentiating between CHW home and neighboring village treatment groups. An obvious pattern of effects, such as larger sizes of coefficient estimates in the home villages, does not emerge. However, these results are not surprising given the relatively modest differences in treatment coverage between the home and neighboring villages. The only statistically significant result — at the 10 percent level — is that the home villages of CHWs experience a 14.2 percentage point differential reduction, relative to control, in improper umbilical cord care.

#### 3.4.2.2 CHW Gender

Table 3.7 shows the results for the estimation differentiating between male and female CHW treatment groups. Again, an obvious pattern of effects does not emerge. For improper umbilical cord care, the male CHW treatment group experiences a 14.5 percentage point differential reduction ( $p < 0.05$ ), relative to control, while the null hypothesis of no effect in the female CHW group cannot be rejected. The female CHW treatment group experiences an

TABLE 3.4: PRE-PROGRAM FALSIFICATION TESTS FOR HOME VS. NEIGHBORING VILLAGE CHW TREATMENT

	-4 to -3 Years	-3 to -2 Years	-2 to -1 Years
<i>Panel A: Four ANC Visits</i>			
Post * Treat-Home	0.120 (0.193)	0.125 (0.123)	-0.012 (0.116)
Post * Treat-Neigh	0.228 (0.210)	-0.035 (0.144)	-0.068 (0.159)
<i>Panel B: Skilled Delivery</i>			
Post * Treat-Home	0.149 (0.181)	0.043 (0.139)	-0.221 (0.160)
Post * Treat-Neigh	-0.125 (0.183)	0.092 (0.140)	-0.079 (0.149)
<i>Panel C: Breastfeed 30 Minutes</i>			
Post * Treat-Home	-0.163 (0.199)	0.080 (0.132)	0.169 (0.127)
Post * Treat-Neigh	-0.038 (0.230)	0.062 (0.154)	0.035 (0.129)
<i>Panel D: Improper Cord Care</i>			
Post * Treat-Home	0.032 (0.195)	-0.184 (0.133)	-0.033 (0.161)
Post * Treat-Neigh	-0.031 (0.174)	-0.133 (0.128)	0.109 (0.135)

*Notes:* This table displays the results of pre-program falsification tests for the heterogeneity analysis based on CHW location, between successive years prior to the baseline survey. Heteroskedasticity-robust standard errors, adjusted for clustering at the village level, are shown in parentheses. \*Significant at 10% level, \*\*Significant at 5% level, \*\*\*Significant at 1% level.

11.9 percentage point differential increase ( $p < 0.10$ ) in breastfeeding within thirty minutes of birth, while the null hypothesis of no effect in the male CHW group cannot be rejected. For ANC visits and skilled delivery, neither of the difference-in-differences estimates is significant.

TABLE 3.5: PRE-PROGRAM FALSIFICATION TESTS FOR MALE VS. FEMALE CHW TREATMENT

	-4 to -3 Years	-3 to -2 Years	-2 to -1 Years
<i>Panel A: Four ANC Visits</i>			
Post * Treat-Male	0.055 (0.192)	0.103 (0.130)	-0.082 (0.134)
Post * Treat-Fem	0.370* (0.196)	-0.049 (0.140)	0.042 (0.145)
<i>Panel B: Skilled Delivery</i>			
Post * Treat-Male	0.114 (0.178)	0.012 (0.129)	0.020 (0.114)
Post * Treat-Fem	-0.075 (0.207)	0.094 (0.162)	-0.447*** (0.155)
<i>Panel C: Breastfeed 30 Minutes</i>			
Post * Treat-Male	-0.295 (0.184)	0.150 (0.129)	0.073 (0.120)
Post * Treat-Fem	0.226 (0.212)	-0.082 (0.148)	0.145 (0.153)
<i>Panel D: Improper Cord Care</i>			
Post * Treat-Male	0.079 (0.179)	-0.164 (0.126)	-0.092 (0.132)
Post * Treat-Fem	-0.110 (0.190)	-0.146 (0.146)	0.250 (0.165)

*Notes:* This table displays the results of pre-program falsification tests for the heterogeneity analysis based on CHW gender, between successive years prior to the baseline survey. Heteroskedasticity-robust standard errors, adjusted for clustering at the village level, are shown in parentheses. \*Significant at 10% level, \*\*Significant at 5% level, \*\*\*Significant at 1% level.

### 3.4.3 Heterogeneity by Respondent Characteristics

Respondent characteristics may also modulate the effects of the intervention. To examine the influence of such characteristics, I estimate the following equation:

$$\begin{aligned}
 Y_{ijt} = & \beta_0 + \beta_1 T_j + \beta_2 POST_t + \beta_3 H_i + \beta_4 (T_j * H_i) \\
 & + \beta_5 (POST_t * H_i) + \beta_6 (POST_t * T_j) + \beta_7 (POST_t * T_j * H_i) + u_{ijt},
 \end{aligned}
 \tag{3.3}$$

where  $H_i$  denotes the characteristic of interest for women  $i$ . Specifically, I consider time to health facility and gestational age.

TABLE 3.6: INTENTION-TO-TREAT EFFECT HETEROGENEITY BASED ON CHW LOCATION

	(1)	(2)	(3)	(4)
	Four ANC Visits	Skilled Delivery	Improper Cord Care	Breastfeed 30 Minutes
CHW Home Village	0.137** (0.058)	0.078 (0.058)	0.117** (0.058)	0.020 (0.054)
CHW Neighboring Village	0.095* (0.054)	-0.062 (0.061)	0.051 (0.058)	-0.034 (0.045)
Post	0.187*** (0.038)	0.085** (0.035)	0.060 (0.047)	0.021 (0.048)
Post * Treatment-Home	-0.020 (0.066)	-0.006 (0.050)	-0.142* (0.072)	-0.008 (0.072)
Post * Treatment-Neigh	-0.096 (0.071)	0.092 (0.068)	-0.032 (0.066)	0.046 (0.066)
Constant	0.493*** (0.037)	0.443*** (0.039)	0.583*** (0.039)	0.686*** (0.032)
Mean Counterfactual (Home)	0.816	0.606	0.760	0.727
Mean Counterfactual (Neighbor)	0.774	0.466	0.694	0.673
Mean Control Post	0.679	0.528	0.643	0.707

Notes: This table displays the results of the intention-to-treat estimation for the heterogeneity analysis based on CHW location. The coefficients of interest correspond to the *Post \* Treatment-Home* and *Post \* Treatment-Neigh* variables. Heteroskedasticity-robust standard errors, adjusted for clustering at the village level, are shown in parentheses. \*Significant at 10% level, \*\*Significant at 5% level, \*\*\*Significant at 1% level.



TABLE 3.7: INTENTION-TO-TREAT EFFECT HETEROGENEITY BASED ON CHW GENDER

	(1)	(2)	(3)	(4)
	Four ANC Visits	Skilled Delivery	Improper Cord Care	Breastfeed 30 Minutes
Male CHW	0.096* (0.052)	0.023 (0.057)	0.121** (0.054)	0.016 (0.049)
Female CHW	0.153** (0.059)	-0.007 (0.061)	0.028 (0.063)	-0.041 (0.055)
Post	0.187*** (0.038)	0.085** (0.035)	0.060 (0.047)	0.021 (0.048)
Post * Treatment-Male	-0.030 (0.066)	0.007 (0.053)	-0.145** (0.070)	-0.043 (0.066)
Post * Treatment-Female	-0.098 (0.073)	0.095 (0.062)	0.002 (0.070)	0.119* (0.070)
Constant	0.493*** (0.037)	0.443*** (0.039)	0.583*** (0.039)	0.686*** (0.032)
Mean Counterfactual (Male)	0.775	0.552	0.764	0.723
Mean Counterfactual (Female)	0.833	0.521	0.671	0.666
Mean Control Post	0.679	0.528	0.643	0.707

Notes: This table displays the results of the intention-to-treat estimation for the heterogeneity analysis based on CHW gender. The coefficients of interest correspond to the *Post \* Treatment-Male* and *Post \* Treatment-Female* variables. Heteroskedasticity-robust standard errors, adjusted for clustering at the village level, are shown in parentheses. \*Significant at 10% level, \*\*Significant at 5% level, \*\*\*Significant at 1% level.

### 3.4.3.1 Time to Health Facility

The unexpected negative, though statistically insignificant, program effect for ANC visits may suggest that pregnant women viewed CHWs as a substitute for ANC. Although the program primarily aimed for the CHWs to address demand-side barriers to maternal health through the provision of information, pregnant women may have viewed the CHWs through a supply-side lens as an extension of health services closer to the household level. If this pathway holds, the result may still conform with the general theory put forth at the outset of this article; to the extent that the CHWs were effective at conveying maternal health information, they may have crowded out health-seeking behavior in the formal healthcare system.

Time to the local health facility may help to elucidate whether such substitution affects ANC visits and possibly other outcomes. Given direct and indirect costs of travel, women who reside further from the health facility may be more likely to substitute CHW visits for ANC, implying that  $\beta_7$  in equation (3.3) is negative. In addition, following from the theoretical discussion in Section 3.1.3, the behavior of women who face longer travel to the local health facility may be more constrained by limited bargaining power, again implying a negative sign for  $\beta_7$ . Table 3.8 shows the results of the estimation. Here, the time to health facility variable is the reported time to the health facility where the woman seeks ANC, centered at the mean of 1.08 hours.<sup>8</sup> The  $\beta_7$  estimate for the ANC regression has the expected sign but, along with all of the other  $\beta_6$  and  $\beta_7$  estimates, is not statistically significant. A clear pattern does not emerge among the estimates from the other regressions.

---

<sup>8</sup>Because some women may give birth at different locations than their ANC facility, the time to the delivery facility may differ. Nevertheless, time to health facility may reflect more generalized levels of access to healthcare.

TABLE 3.8: INTENTION-TO-TREAT EFFECT HETEROGENEITY BASED ON HOURS TO FACILITY

	(1)	(2)	(3)	(4)
	Four ANC Visits	Skilled Delivery	Improper Cord Care	Breastfeed 30 Minutes
Treatment	0.114** (0.047)	-0.003 (0.048)	0.087* (0.049)	-0.012 (0.043)
Post	0.191*** (0.037)	0.092** (0.036)	0.046 (0.048)	0.030 (0.047)
Hours to Facility	-0.010 (0.035)	-0.072* (0.041)	0.037 (0.038)	-0.028 (0.040)
Treatment * Hours to Facility	-0.032 (0.053)	-0.052 (0.064)	0.001 (0.053)	-0.046 (0.055)
Post * Hours to Facility	-0.026 (0.053)	-0.028 (0.040)	0.034 (0.048)	-0.051 (0.050)
Post * Treatment	-0.054 (0.057)	0.039 (0.049)	-0.075 (0.063)	0.014 (0.060)
Post * Treatment * Hours to Facility	-0.010 (0.073)	0.030 (0.067)	0.007 (0.066)	0.079 (0.068)
Constant	0.492*** (0.036)	0.454*** (0.037)	0.590*** (0.039)	0.686*** (0.032)

Notes: This table displays the results of the intention-to-treat estimation for the heterogeneity analysis based on time to health facility. The coefficient of interest corresponds to the  $Post * Treatment * HourstoFacility$  variable. Heteroskedasticity-robust standard errors, adjusted for clustering at the village level, are shown in parentheses. \*Significant at 10% level, \*\*Significant at 5% level, \*\*\*Significant at 1% level.

### 3.4.3.2 Gestational Age

Women with lower gestational age (recorded in weeks) at the time of their first ANC visit may experience greater exposure to the intervention, since CHWs will in principle have more opportunities to visit these women. If such a pathway is active, the sign of  $\beta_7$  should oppose the sign of  $\beta_6$ ; that is, the further along in pregnancy a woman is at the time of her first ANC visit, the lesser the effect of treatment. Table 3.9 shows the results of the estimation, in which gestational age has been mean-centered at 18.99 weeks of pregnancy. While none of the program effect estimates are statistically significant, they all conform to the hypothesis that  $\beta_7$  should oppose  $\beta_6$ . For the skilled delivery, improper cord care, and breastfeeding regressions, the  $\beta_7$  estimates are relatively small. In the ANC visits regression, the effect of gestational age alone, absent treatment, is statistically significant and logically negative; almost mechanically, greater gestational age makes it more difficult to complete all four ANC visits. As in the main specification, the primary difference-in-differences coefficient estimate ( $\hat{\beta}_6$ ) for ANC visits remains unexpectedly negative. By a similar token, the estimate of  $\beta_7$  is surprisingly positive. The estimate implies that, for women who attend their first ANC visit at eight weeks beyond the mean gestational age, treatment increases the likelihood of ANC visit completion by 11.2 percentage points. If this effect is true, and if the ANC-CHW substitution hypothesis described above holds, the positive coefficient may reflect reduced substitution among women with lower exposure to the intervention. Alternatively, if no ANC-CHW substitution is present, then the effect could suggest that CHW visits help to ‘catch up’ women who face more difficult pregnancy situations due to late presentation for ANC. Alongside these possibilities, the failure to detect a statistically significant effect, across both the primary and heterogeneity estimates, may arise from insufficient power.

TABLE 3.9: INTENTION-TO-TREAT EFFECT HETEROGENEITY BASED ON GESTATIONAL AGE

	(1)	(2)	(3)	(4)
	Four ANC Visits	Skilled Delivery	Improper Cord Care	Breastfeed 30 Minutes
Treatment	0.115** (0.046)	0.010 (0.049)	0.079 (0.051)	-0.012 (0.042)
Post	0.189*** (0.035)	0.077** (0.035)	0.060 (0.046)	0.021 (0.049)
Gestational Age	-0.012** (0.005)	-0.007 (0.004)	-0.004 (0.005)	0.004 (0.005)
Treatment * Gestational Age	-0.001 (0.007)	0.008 (0.006)	-0.001 (0.006)	-0.004 (0.006)
Post * Gestational Age	-0.011* (0.006)	0.001 (0.006)	0.005 (0.007)	0.003 (0.006)
Post * Treatment	-0.062 (0.054)	0.043 (0.047)	-0.086 (0.060)	0.025 (0.060)
Post * Treatment * Gestational Age	0.014 (0.008)	-0.004 (0.008)	0.006 (0.009)	-0.007 (0.008)
Constant	0.493*** (0.037)	0.450*** (0.037)	0.588*** (0.040)	0.687*** (0.031)

Notes: This table displays the results of the intention-to-treat estimation for the heterogeneity analysis based on gestational age. The coefficient of interest corresponds to the  $Post * Treatment * Gestational Age$  variable. Heteroskedasticity-robust standard errors, adjusted for clustering at the village level, are shown in parentheses. \* Significant at 10% level, \*\* Significant at 5% level, \*\*\* Significant at 1% level.

### 3.4.4 Primigravida Women

As noted above, the difference-in-differences estimation procedure excludes primigravida women due to the absence of baseline outcome variables for these individuals. To characterize the program's effects on primigravida women, I estimate the following specification:

$$Y_{ij}^{post} = \delta_0 + \delta_1 T_j + \delta_2 PG_{ij} + \delta_3 (T_j * PG_{ij}) + \delta_4 Y_{ij}^{pre} + \nu_{ij}, \quad (3.4)$$

where  $Y_{ij}^{post}$  is the post-program outcome for woman  $i$  in village  $j$ ;  $Y_{ij}^{pre}$  is the pre-program outcome, which I set to zero for all primigravida women;  $PG_{ij}$  is an indicator variable equal to 1 for primigravida women; and  $\nu_{ij}$  is an error term. In contrast to the difference-in-differences specification, the left-hand side of this regression equation is the post-program outcome variable level, which changes the interpretation of the coefficient estimates. The sum  $\delta_1 + \delta_3$  gives the difference in the post-program outcome for primigravida women in the treatment group compared to primigravida women in the control group. Note, however, that the specification does not difference out the program-group fixed effects, which means that they will be absorbed by the  $\delta_1$  and  $\delta_3$  coefficients. As a result, we must be cautious in interpreting the results, as they are not causally identified.

Table 3.10 displays the results of the estimation. Primigravida women in the treatment group exhibit little difference in antenatal visit completion compared to those in the control group. For skilled delivery, primigravida women show a 9.6 percentage point reduction associated with treatment. Improper umbilical cord care is also notable, with primigravida women in the treatment group exhibiting an 11.6 percentage point increase in the likelihood of this practice. Apart from treatment differences, the regressions also demonstrate that being primigravida itself tends to be a strong predictor of the outcomes.

TABLE 3.10: PROGRAM EFFECTS FOR PRIMIGRAVIDA WOMEN.

	(1)	(2)	(3)	(4)
	Four ANC Visits	Skilled Delivery	Improper Cord Care	Breastfeed 30 Minutes
Treatment	0.036 (0.045)	0.050 (0.044)	-0.021 (0.048)	0.012 (0.048)
Primigravida (PG)	0.177** (0.070)	0.526*** (0.058)	0.143* (0.081)	-0.059 (0.095)
Treatment * PG	-0.035 (0.099)	-0.146* (0.087)	0.137 (0.097)	0.058 (0.104)
Previous Four ANC	0.225*** (0.037)			
Previous Skilled Delivery		0.403*** (0.034)		
Previous Improper Cord Care			0.193*** (0.043)	
Previous Breastfeed 30 Minutes				0.082 (0.054)
Constant	0.568*** (0.039)	0.347*** (0.042)	0.530*** (0.043)	0.650*** (0.058)
Mean Dep. Var.	0.725	0.600	0.649	0.697

Notes: This table displays the results of the estimation of program effects for primigravida women. Heteroskedasticity-robust standard errors, adjusted for clustering at the village level, are shown in parentheses. \*Significant at 10% level, \*\*Significant at 5% level, \*\*\*Significant at 1% level.

### 3.4.5 Program Feedback

Given the absence of any striking results, program feedback solicited from respondents who received one or more CHW visits may shed light regarding the CHW intervention. Table 3.11 displays a list of feedback questions or responses, with the proportion of subjects answering “yes” reported in the right-hand column. On the whole, respondents gave positive feedback on the CHWs. Forty-five percent of women report that the CHWs influenced the number of ANC visits they attended, while 37.7 percent report that the CHWs influenced their delivery location. Far fewer report influential effects on postnatal care, which is consistent with relatively low postnatal care levels that persist in such communities. Ninety-eight percent of women report that they liked when the CHW visited, and 82.5 percent rated the overall performance of the CHW as either good or very good, the two highest ratings possible. However, only 29.8 percent of respondents reported that the CHW talked to their husbands. When asked an open-ended question soliciting suggestions for future program improvements, 53.6 percent and 59.5 percent of women independently requested more frequent CHW visits and the distribution of more information, respectively, again seeming to indicate positive perceptions of the intervention. Despite the positive nature of the feedback, we must take respondent feedback with a grain of salt, since respondent perceptions may diverge from tangible, recorded outcomes. Nevertheless, such feedback may be an important signal of welfare benefits not captured by the analysis above. At the same time, with only 59.2 percent of women in the treatment group reporting one or more visits from a CHW, and then only a fraction of these women reporting that their health behaviors were influenced by the CHW, the treatment quality may have been too diluted to produce an observable program effect.



TABLE 3.11: RECIPIENT ASSESSMENT OF CHW VISITS

	Mean
Talked about something you did not know?	0.353
Influenced # ANC visits?	0.450
Influenced ANC location?	0.448
Influenced delivery location?	0.377
Influenced # postnatal visits?	0.120
Influenced postnatal care location?	0.135
Liked when CHW came?	0.980
Talked to husband?	0.298
More frequent visits?	0.536
Distribute more information?	0.595
Overall performance good or very good	0.825

*Notes.* This table displays binary variable averages of responses to feedback questions asked of program recipients.

### 3.4.6 Pre-Post Changes

Despite the absence of clear program effects, the coefficient on  $POST_t$  in Table 3.3 points to significant changes over time for antenatal visit completion and skilled delivery in both the treatment and control groups. This result suggests that the single difference across pre- and post-program periods for the entire sample, rather than the double difference distinguishing treatment effects, may be important to examine. Table 3.12 compares the pre- and post-program means for the four outcomes in the entire sample. The four ANC visits and skilled delivery outcomes show highly statistically significant 27.3 percent and 24.2 percent increases over baseline levels, respectively. As further detailed in the conclusion, several different explanations — such as PHP’s concurrent health facility improvements, spillovers from the CHWs, and background health trends — may underlie these changes. Regardless of the specific explanation, significant background changes over time likely render the estimation of the causal effect of the CHW home visits more challenging.

TABLE 3.12: PRE- TO POST-INTERVENTION CHANGES FOR ENTIRE SAMPLE

	(1)	(2)	(3)	(4)	(5)
	Mean Pre	Mean Post	Difference	Percent Change	<i>p</i> -value
Four ANC Visits	0.562	0.716	0.154	27.3	0.000
Skilled Delivery	0.450	0.559	0.109	24.2	0.000
Improper Cord Care	0.635	0.641	0.006	0.9	0.845
Breastfeed 30 Minutes	0.683	0.714	0.031	4.6	0.282

*Notes.* This table displays pre-intervention and post-intervention outcome means for the overall sample, along with *p*-values for difference-in-means tests across these periods. All tests use heteroskedasticity-robust standard errors adjusted for clustering at the village level.

### 3.4.7 Predictors of Outcomes

In an effort to better understand the determinants of program outcomes, I perform a regression of post-program outcomes on a selection of covariates representing socioeconomic status, health history, access to healthcare, and health practices. All of the predictors are baseline covariates, with the exception of number of symptoms during pregnancy and husband deciding delivery location, which are reported at endline. While not causal in nature, the regression controls for treatment group status and provides some insight into the determinants of outcomes. Table 3.13 displays the results.

Most notably, women whose husbands decide their delivery locations experience a highly statistically significant 24.9 percentage point increase in the probability of skilled delivery. Thus, conditional on a woman’s husband making the delivery decision, male partner preferences appear to favor the welfare of women. However, the broader interpretation of this observation is unclear. What are the implications for women’s bargaining power if male partners, rather than women, make such decisions, albeit in the interest of the woman’s welfare? If the bargaining power of women remains low, a cultural explanation may prevail: feeling the need to demonstrate their responsibility to care for their families, men may assert themselves by taking such decisions into their own hands. The finding also leaves open questions about male partners who do not make the delivery decision. What is the role of bargaining power under these circumstances, and what differentiates those partners who

make the delivery decision from those who do not?

Respondent education and asset index are also strongly positively correlated with delivery under skilled supervision. For each additional pregnancy in their history, women experience a 2.4 percentage point drop in the probability of skilled delivery. This association could be explained by age differences in health behaviors or by negative learning over time that discourages women from returning to the local health facilities. The negative association of number of pregnancies with improper cord care is subject to similar interpretation, although in this case the learning effect would be positive. Respondents exhibit an 8.1 percentage point reduction in the probability of skilled delivery with each additional hour of travel required to the local health facility, confirming that travel time contributes significantly to health decision-making. Limited access to healthcare also increases the likelihood of improper cord care, perhaps by reducing a woman's exposure to opportunities for health education. Nevertheless, the correlation between hours to the local health facility and ANC visit completion, while negative, is not statistically significant. The number of symptoms a woman experiences during pregnancy is also associated with a statistically significant increase in the likelihood of skilled delivery. While this correlation may reflect risk-averse behavior in response to the detection of health risks, it could also simply reflect more generalized risk aversion, embodied by women who have more health-conscious dispositions. Finally, in contrast to the three other outcomes, breastfeeding within thirty minutes of birth has no significant predictors.

## **3.5 Discussion**

### **3.5.1 Treatment Effects**

This article provides evidence on the maternal health outcomes of a CHW home visit program, a type of intervention that has become increasingly emphasized in the healthcare systems of many low- and middle-income countries in recent years. Despite the excitement around such interventions, I observe mixed results. I find no statistically significant program effects on completing four antenatal care visits, giving birth with skilled attendance, breast-

TABLE 3.13: PREDICTORS OF POST-PROGRAM OUTCOMES

	(1)	(2)	(3)	(4)
	Four ANC Visits	Skilled Delivery	Improper Cord Care	Breastfeed 30 Minutes
Treatment	0.039 (0.035)	0.012 (0.038)	0.035 (0.039)	0.006 (0.040)
Respondent Education	0.002 (0.005)	0.018*** (0.006)	-0.007 (0.007)	0.011 (0.007)
Asset Index	0.000 (0.010)	0.037*** (0.010)	0.006 (0.011)	0.013 (0.011)
# Pregnancies	-0.002 (0.007)	-0.024*** (0.007)	-0.021** (0.009)	0.006 (0.007)
Hours to Facility	-0.034 (0.030)	-0.081*** (0.022)	0.059*** (0.022)	-0.038 (0.030)
Previous Pregnancy Death	0.046 (0.044)	0.020 (0.058)	0.037 (0.047)	0.036 (0.050)
Gestational Age First ANC	-0.016*** (0.003)	-0.001 (0.003)	0.005 (0.003)	0.003 (0.002)
Owens Mosquito Bednet	0.052 (0.046)	0.036 (0.047)	-0.124*** (0.045)	0.025 (0.048)
Husband Makes Health Decisions	-0.013 (0.030)	-0.046 (0.033)	0.020 (0.037)	0.015 (0.031)
# Symptoms	0.002 (0.007)	0.018** (0.008)	0.024*** (0.008)	-0.014 (0.008)
Husband Decided Delivery Location		0.249*** (0.035)		
Mean Dep. Var.	0.725	0.600	0.649	0.697

*Notes:* This table displays regression results for the predictors of post-program outcomes. Heteroskedasticity-robust standard errors, adjusted for clustering at the village level, are shown in parentheses. Please see Appendix C.1.1 for further detail on how missing values are treated. \*Significant at 10% level, \*\*Significant at 5% level, \*\*\*Significant at 1% level.

feeding within thirty minutes of birth, and engaging in improper umbilical cord care, either through the primary ITT estimation or through the heterogeneity analyses. Although not statistically significant, the estimated ITT effects for skilled delivery, breastfeeding within thirty minutes, and improper umbilical cord care have the expected signs. However, the estimated ITT effect for antenatal visit completion is negative, moving in the unexpected direction.

If we believe that this effect signals a meaningful change, one interpretation may be that pregnant women view CHWs as a substitute for antenatal care. Based on theory, we might expect that women who reside further from the health facility — and therefore experience a greater cost of seeking care — exhibit stronger substitution between CHWs and antenatal care. However, the analysis of heterogeneous effects based on time to health facility does not conclusively confirm this hypothesis; although the treatment effect becomes increasingly negative with greater time to the health facility, the estimate is small and is not statistically significant. At the same time, the analysis of heterogeneity by gestational age at first ANC visit provides suggestive evidence that treatment effects decrease with increasing gestational age, which should indicate reduced exposure to the intervention. This analysis also shows a pattern of enhanced substitution between CHWs and antenatal care for women with greater exposure to the intervention. Finally, across all of the results, it is important to keep in mind that the failure to reject some of the null hypotheses tested herein may arise from insufficient statistical power.

For the improper cord care and breastfeed within thirty minutes outcomes, I find a few notable heterogeneous effects. Women visited by male CHWs show a 14.5 percentage point reduction in improper cord care, while women visited by female CHWs show an 11.9 percentage point increase in breastfeeding within thirty minutes of birth. In addition, women who reside in a CHW's home village — where treatment intensity may be greater — show a 14.2 percentage point reduction in improper cord care. However, the broader explanation for these results, particularly in the context of all of the outcomes together, remains unclear.

### 3.5.2 Pre-Post Changes

While the ITT effects appear to be weak, the pre-post changes in antenatal visit completion and skilled delivery for the entire sample are striking. The following possibilities may account for these trends: (1) concurrent health facility improvements, (2) spillover effects of the CHWs, and/or (3) background health trends. First, the changes may have been driven by the first component of PHP's Omukazi Namagara Program, which included quality of care improvements at the local government health centers that are accessible to all women in the study area. Women and households may have been far more responsive to the quality of care improvements than to the CHW home visits. Such a finding would shine an important light on the influence of supply-side factors as determinants of health decision-making, in contrast to the more common view of demand-side frictions that impede neoclassical behavior.

Attributing the pre-post change to the health service strengthening alone is complicated by the abovementioned information campaign that PHP carried out in all 106 villages at the start of the intervention. However, the facility improvements likely play a prominent role. As described above, the substantial attrition in the survey appears to be largely due to women from localities outside the study area who sought services at the health facilities. Field reports indicate that these women opted to visit these health facilities because they had learned of the improved quality of care. As it may be unlikely for news about the facility improvements to spread very quickly from the experiences of individual patients — particularly given the low flow of patients relative to the whole population — it is entirely possible that the information campaign and the facility improvements together produced an interactive effect on healthcare-seeking behavior. Ultimately, such a result would still point to the responsiveness of women to supply-side factors. It is unlikely that the information campaign alone, and in particular its demand-side messages, having been carried out at the start of the program and at a single point in time, would have had a sustained impact on outcomes for the extended duration of the program.

An alternative explanation for the large pre-post changes is that the CHW component of the program produced treatment-to-control spillover effects, possibly in interaction with the health facility improvements. In this case, the CHW visits may have had a substantial

impact on health, but the effect would be difficult to causally detect. As the CHWs are unlikely to have made home visits in the control villages<sup>9</sup>, such spillover effects would have operated through the spread of information disseminated through the CHW visits. The presence of spillovers alone could explain the pre-post changes across the entire sample and would indicate an effect of the CHW home visits; however, it is also possible that the information spread through spillovers included information regarding the health facility improvements and — similar to the interactive effect described above — impacted healthcare-seeking behavior between the pre- and post-intervention periods through a complementary mechanism of both facility improvements and CHW information dissemination.

A third possibility is that the pre-post changes reflect health trends already underway in Uganda and/or the study localities in particular. According to data from the Demographics and Health Survey, the completion of four antenatal care visits in rural Uganda increased from 45.8 to 58.3 percent between 2011 and 2016, and deliveries in a health facility in southwestern Uganda increased from 40.3 to 70.6 percent over the same period (UBOS and ICF 2012; 2018). In addition, the pre-program trend for antenatal care completion appears to show a clear upwards path; the trend for skilled delivery is less evident. Related to the background trends possibility, the pre-post changes could also be an artefact of the construction of baseline outcomes. Since the baseline outcomes stretch as far as five years prior to the intervention, the influence of background health trends on the before-after change may be heightened.

### 3.5.3 CHW Management and Performance

If the intervention truly produced no impact, it is important to explore why. The 59.2 percent CHW coverage rate reported by women in the treatment group stands out as a key

---

<sup>9</sup>With regard to the team of CHWs working on PHP's program, the organization only compensated them for home visits within their two assigned villages. Given that the CHWs did not even fully cover the women in the treatment villages, they would have had little incentive to conduct extraneous home visits in outside villages, which almost certainly would have been more geographically distant. Indeed, a key empirical observation from this study, discussed further in the conclusion, relates to the importance of management and incentive structures motivating CHW performance. In addition, as mentioned above in regard to the parallel trends assumption, I am not aware of any other major programs or policy changes related to maternal health, and in particular involving the PHP CHWs, that occurred in Kashongi and/or Kitura during the study period.

explanation. Moreover, the impact of CHWs depends jointly on performance along both the extensive and intensive margins. On the extensive margin, the CHW decides whether to visit a woman or not; on the intensive margin, the CHW performs according to dimensions such as those featured in Table 3.11. Given this joint determination of performance, the “effective coverage” may have been substantially below 59.2 percent, reducing the quality of treatment.

CHW performance is a matter of its own that warrants attention, given that it is in the interest of managers to maximize the performance of CHWs. While I do not have any quantitative data from the CHWs themselves to determine the reasons for partial coverage, there are several possible explanations based on observations from the field. First, because many of the CHWs tend to be more educated members of the community, sometimes with additional sources of income, PHP’s payments to the CHWs may not sufficiently compensate for the opportunity cost of their time. Second, as described above, the CHWs may have had trouble identifying pregnant women, particularly in their neighboring villages where they may not have known all of the households. Third, because PHP compensated CHWs based on the records of home visits that they submitted, some CHWs may have been incentivized to fabricate the records without actually visiting all of the women. While PHP instituted monitoring mechanisms to help mitigate this possibility, such monitoring was relatively lax over the course of the study period. Lastly, it is possible that CHWs selectively chose to visit some women but not others, perhaps based on social networks or perceived need.

Overall, interventions of this nature may require substantial managerial resources and oversight to achieve a high rate of coverage. The success of CHW programs at a policy level may hinge on the monitoring mechanisms and performance incentives that programs implement. However, these systems are not free; the benefits of higher coverage must be weighed against the accompanying administrative costs. Future research should focus on the supply side by evaluating the monitoring and incentive schemes that programs can implement to achieve optimal CHW effort.



## References

- Adam, Mary B, Maria Dillmann, Mei-kuang Chen, Simon Mbugua, Joram Ndung'u, Priscilla Mumbi, Eunice Waweru, and Peter Meissner. 2014. "Improving Maternal and Newborn Health: Effectiveness of a Community Health Worker Program in Rural Kenya." *PloS One* 9 (8): e104027.
- Akinyi, Caroline, Jeremie Nzanzu, and Dan Kaseje. 2015. "Effectiveness of Community Health Workers in Promotion of Maternal Health Services in Butere District, Rural Western Kenya." *Universal Journal of Medical Science* 3 (1): 11–18.
- Allendorf, Keera. 2007. "Couples' Reports of Women's Autonomy and Health-care Use in Nepal." *Studies in Family Planning* 38 (1): 35–46.
- Ashraf, Nava, Erica Field, and Jean Lee. 2014. "Household Bargaining and Excess Fertility: An Experimental Study in Zambia." *American Economic Review* 104 (7): 2210–2237.
- Becker, Stan, Fannie Fonseca-Becker, and Catherine Schenck-Yglesias. 2006. "Husbands' and Wives' Reports of Women's Decision-making Power in Western Guatemala and Their Effects on Preventive Health Behaviors." *Social Science & Medicine* 62 (9): 2313–2326.
- Beegle, Kathleen, Elizabeth Frankenberg, and Duncan Thomas. 2001. "Bargaining Power within Couples and Use of Prenatal and Delivery Care in Indonesia." *Studies in Family Planning* 32 (2): 130–146.
- Cairncross, Sandy, Kathleen Shordt, Suma Zacharia, and Beena Kumari Govindan. 2005. "What Causes Sustainable Changes in Hygiene Behaviour? A Cross-sectional Study from Kerala, India." *Social Science & Medicine* 61 (10): 2212–2220.
- Cohen, Jessica, Pascaline Dupas, and Simone Schaner. 2015. "Price Subsidies, Diagnostic Tests, and Targeting of Malaria Treatment: Evidence from a Randomized Controlled Trial." *American Economic Review* 105 (2): 609–645.
- Danforth, EJ, ME Kruk, PC Rockers, G Mbaruku, and S Galea. 2009. "Household Decision-Making about Delivery in Health Facilities: Evidence from Tanzania." *Journal of Health, Population and Nutrition*: 696–703.
- Doss, Cheryl. 2013. "Intrahousehold Bargaining and Resource Allocation in Developing Countries." *World Bank Research Observer* 28 (1): 52–78.
- Dupas, Pascaline. 2011a. "Do Teenagers Respond to HIV Risk Information? Evidence from a Field Experiment in Kenya." *American Economic Journal: Applied Economics* 3 (1): 1–34.
- . 2011b. "Health Behavior in Developing Countries." *Annual Review of Economics* 3 (1): 425–449.
- Filmer, Deon, and Lant H Pritchett. 2001. "Estimating Wealth Effects without Expenditure Data—or Tears: An Application to Educational Enrollments in States of India." *Demography* 38 (1): 115–132.
- Geldsetzer, Pascal, et al. 2019. "Community Health Workers to Improve Uptake of Maternal Healthcare Services: A Cluster-Randomized Pragmatic Trial in Dar es Salaam, Tanzania." *PLoS Medicine* 16 (3): e1002768.

- Hou, Xiaohui, and Ning Ma. 2013. "The Effect of Women's Decision-Making Power on Maternal Health Services Uptake: Evidence from Pakistan." *Health Policy and Planning* 28 (2): 176–184.
- Jalan, Jyotsna, and E Somanathan. 2008. "The Importance of Being Informed: Experimental Evidence on Demand for Environmental Quality." *Journal of Development Economics* 87 (1): 14–28.
- Kirkwood, Betty R, et al. 2013. "Effect of the Newhints Home-Visits Intervention on Neonatal Mortality Rate and Care Practices in Ghana: A Cluster Randomised Controlled Trial." *The Lancet* 381 (9884): 2184–2192.
- Madajewicz, Malgosia, Alexander Pfaff, Alexander Van Geen, Joseph Graziano, Iftikhar Hussein, Hasina Momotaj, Roksana Sylvi, and Habibul Ahsan. 2007. "Can Information Alone Change Behavior? Response to Arsenic Contamination of Groundwater in Bangladesh." *Journal of Development Economics* 84 (2): 731–754.
- Maitra, Pushkar. 2004. "Parental Bargaining, Health Inputs and Child Mortality in India." *Journal of Health Economics* 23 (2): 259–291.
- Peters, Christina. 2011. "Effects of Family Planning and Health Services on Women's Welfare: Evidence on Dowries and Intra-household Bargaining in Bangladesh." *Review of Economics of the Household* 9 (3): 327–348.
- Rhee, Michelle, Mahamadou Sissoko, Sharon Perry, Willi McFarland, Julie Parsonnet, and Ogobara Doumbo. 2005. "Use of insecticide-treated Nets (ITNs) Following a Malaria Education Intervention in Piron, Mali: A Control Trial with Systematic Allocation of Households." *Malaria Journal* 4 (1): 1.
- Singh, Prabhjot, and Jeffrey D Sachs. 2013. "1 Million Community Health Workers in Sub-Saharan Africa by 2015." *Lancet* 382 (9889): 363–365.
- UBOS and ICF. 2012. *Uganda Demographic and Health Survey 2011*. Country Report. Kampala, Uganda and Calverton, Maryland: Uganda Bureau of Statistics and ICF International Inc.
- Waiswa, Peter, et al. 2015. "Effect of the Uganda Newborn Study on Care-Seeking and Care Practices: A Cluster-Randomised Controlled Trial." *Global Health Action* 8 (1): 24584.
- WHO. 2012. *WHO Recommendations: Optimizing Health Worker Roles to Improve Access to Key Maternal and Newborn Health Interventions through Task Shifting*. Tech. rep. World Health Organization.
- WHO, UNICEF, UNFPA, World Bank Group, and United Nations Population Division. 2019. *Trends in Maternal Mortality: 2000 to 2017: Estimates by WHO, UNICEF, UNFPA, World Bank Group and the United Nations Population Division*. Tech. rep. World Health Organization.

## CHAPTER 4

### Disparity Breeds Disparity: The Impact of Racial Inequalities in Incarceration on Educational Choice

#### 4.1 Introduction

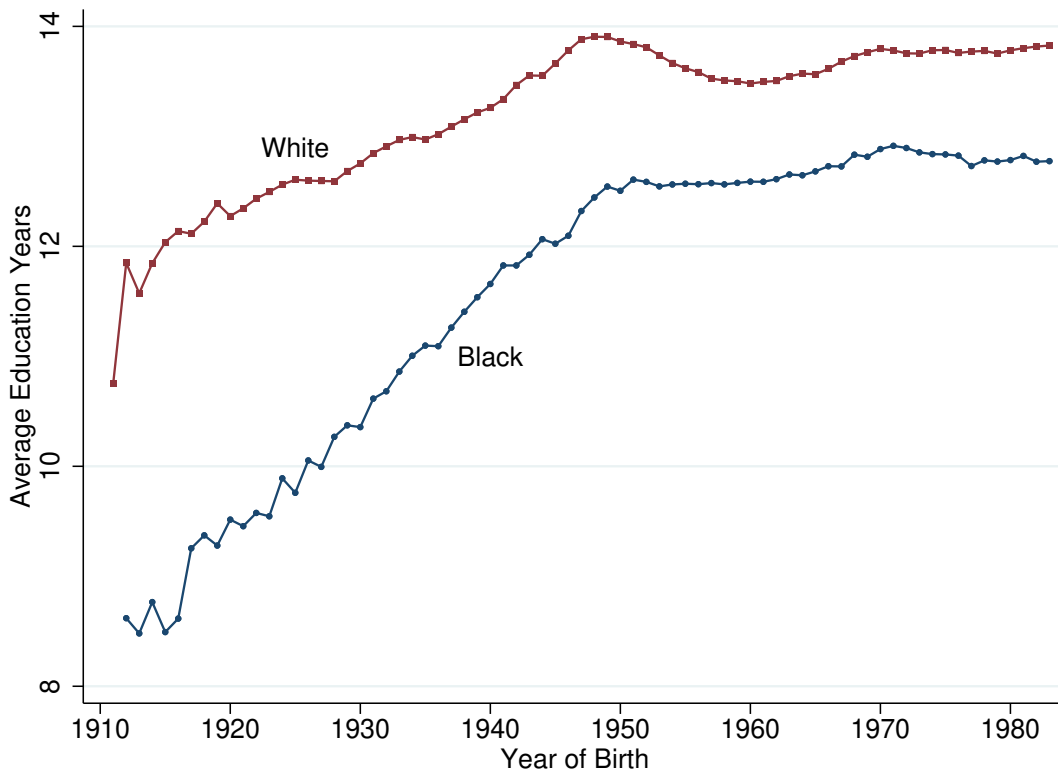
Mass incarceration produces major, direct social consequences for many Black individuals and families, from deleterious health effects to lost economic productivity. But inequalities in the criminal justice system may also operate insidiously: by burdening young Black men with substantial *expectations of future incarceration* and thereby altering their life decisions before they even experience any imprisonment. This pathway may play a critical role in driving other major racial disparities — for example, in education. In fact, this article demonstrates that this pathway explains most, if not all, of the current education gap between Black and White men — a significant finding given that Black men obtain 12.8 years of education on average, while White men invest an average of 13.8 years in school, according to recent data.<sup>1</sup> As shown in Figure 4.1, such a substantial gap in educational attainment has persisted for decades, with relatively little change since Black male birth cohorts from the early 1950s stopped catching up to White men.

How does the incarceration-to-education pathway operate? It is grounded in basic economic intuition: an individual who expects to be incarcerated — and thus removed from the labor market in prime-earning years — has diminished incentives to go to school. Incarceration effectively reduces the time over which the benefits of educational investments can be realized. This loss of time has increased in recent decades in the U.S., especially for the average male, and even more so for the average Black male, whose point-in-time

---

<sup>1</sup>These estimates are based on pooled 2010-2014 *American Community Survey* data on the educational attainment of males between the ages of 30 and 39 (inclusive), which we take to be the age bracket that has most recently completed its schooling.

FIGURE 4.1: MALE EDUCATION TRENDS BY RACE



*Notes:* The trends shown in this figure consist of average educational attainment of 30-and-over White and Black males by year of birth, computed from pooled 2006-2014 *American Community Survey* data, for birth cohorts ranging from 1911 to 1983. The included birth cohorts end with 1983 because this year corresponds to the most recent cohort whose members are at least 30 years of age in 2014 (the final year of the data), and who can therefore be presumed to have finished schooling. However, as the assumption that individuals 30 years and older have finished schooling may not always hold, the computation of average educational attainment for the most recent birth cohorts in the graph may not be as accurate as that for earlier birth cohorts. Correspondingly, the educational attainment shown for the most recent birth years may not reflect true, final education outcomes.

incarceration rate is 6.9 percent on average between the ages of 18 and 60. This number is roughly five-fold that of White males, which is, in turn, almost five times larger than the rate for White females.<sup>2</sup> Faced with significant incarceration risks at working ages, Black men invest substantially less time in education.

As motivation for this approach, consider Figure 4.2. Panel A shows the disparate incarceration rates between Black and White men. Panel B applies these incarceration rates to the analysis of survival over the life course between the two groups. The solid and long-dashed

<sup>2</sup>These estimates are based on 2014 *American Community Survey* data on institutionalization. For the age range of 18 to 60, such data almost entirely reflect incarceration. See Section 4.3.5 and Appendix D.2.1 for further detail on how we work with the institutionalization data.

lines show standard demographic survival curves, conditional on surviving to age 18<sup>3</sup>, for non-Hispanic White and Black men, respectively. These curves show the substantial baseline disparity in mortality risk over the life course between these two groups. But, like mortality, incarceration leads to lost time by eliminating years of labor force participation. The short- and medium-dashed lines account for this effect by adjusting the standard mortality survival curves for incarceration risk over the life course; specifically, they show the probability of being both alive and not incarcerated at a given age.

This adjustment elucidates the significant additional disparity between Black and White men created by incarceration risk. In the 20s, 30s, and 40s — prime working years — Black men have a far lower probability of pseudo-survival, almost entirely due to incarceration. From the age of 18 (grade 12), when many individuals decide whether to continue with school, Black males can expect to live an additional 52.56 incarceration-adjusted years, roughly six years less than their White male counterparts.

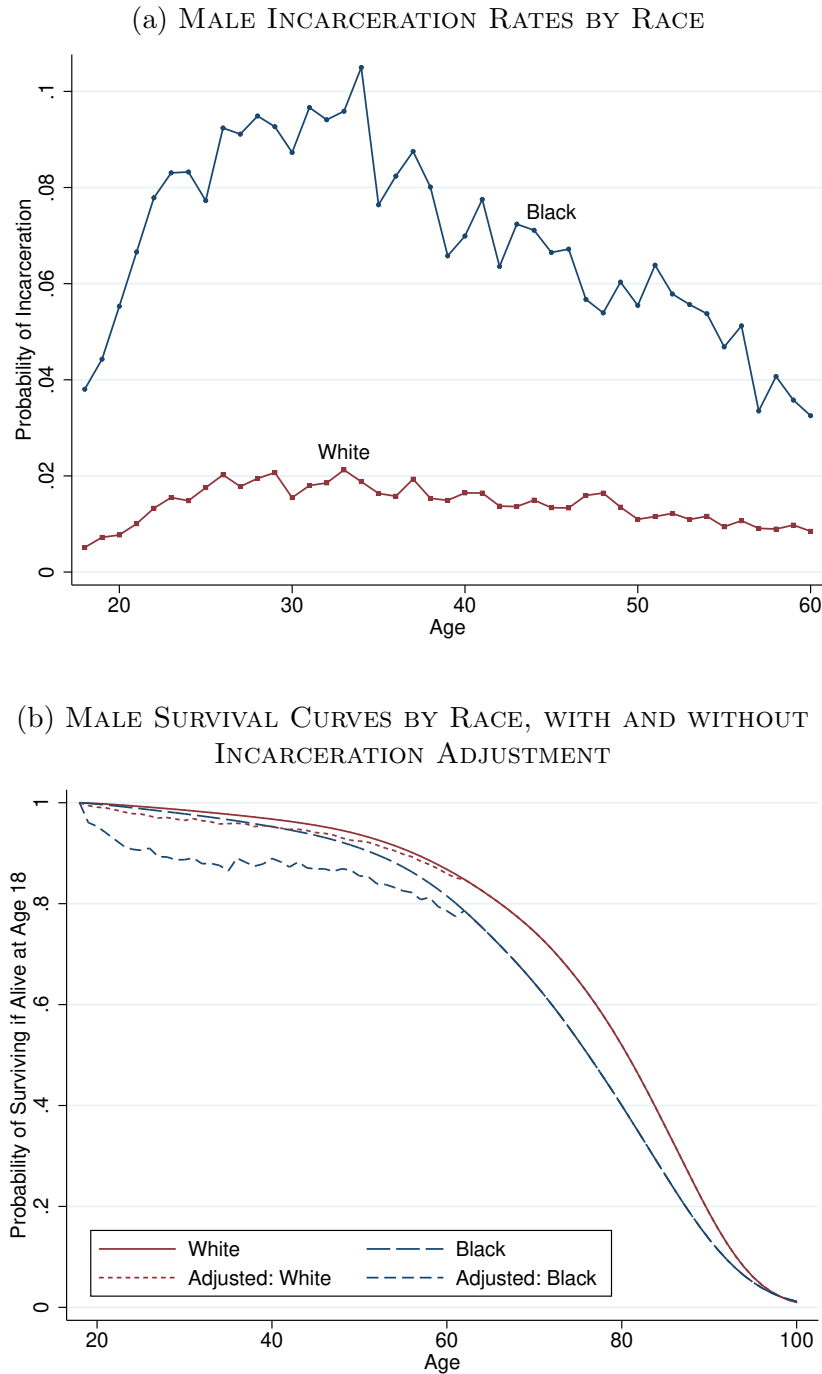
Building on these observations, we ask the following question in this article: to what extent can incarceration and mortality risk — by depressing expected labor force participation and thus earnings over the lifetime — explain education disparities between Black and White men? In particular, if White men were hypothetically given the same incarceration and mortality rates as Black men, by how many years would they reduce their educational investment? We hypothesize that incarceration and mortality disparities are responsible for a substantial portion of the education gap, giving prominence to the incarceration-to-education pathway relative to the reverse pathway.

To examine this hypothesis, we treat incarceration and mortality risk as constraints and analyze the education disparity between Black and White men as a rational outcome of optimizing behavior in the face of these constraints. Specifically, we employ an adaptation of the Mincer model of education — a canonical economic model with more than fifty years of literature behind it — to characterize the educational choice of an individual facing incarceration and mortality risk. We map this theoretical model to data from the *American Community*

---

<sup>3</sup>The survival curve, often denoted  $\ell(t)$  by demographers, is meant to measure the probability of being alive at a given age. The conditional survival curve for some age  $a$  is given by  $\ell_a(t) = \frac{\ell(t)}{\ell(a)}$ , where  $\ell(a)$  is the probability of surviving to age  $a$ . Please see Appendix D.2.2 for more information on the formulation of conditional survival curves.

FIGURE 4.2: MALE SURVIVAL AND INCARCERATION BY RACE



*Notes:* The incarceration rates shown in Panel A, which use institutionalization data to represent incarceration between the ages of 18 and 60, are based on 2014 ACS data and consist of average institutionalization-by-age probabilities for Black and White men. The survival curves shown in Panel B are based on 2014 data from the *National Vital Statistics Reports* and are conditional on surviving to age 18. The solid and long-dash lines represent standard mortality-based survival curves, while the short-dash and medium-dash lines represent survival adjusted for the incarceration rates from Panel A. We perform this adjustment using the following equation:  $\ell^{adj}(t) = \ell(t) - m(t-1) \cdot \ell(t-1)$ , where  $\ell(t)$  denotes survival and  $m(t)$  denotes incarceration. Because the adjustment uses institutionalization data, we institute an age cutoff of 60, above which we assume that observed institutionalization no longer represents incarceration and that the incarceration rate  $m$  becomes zero. Therefore, above the cutoff, the unadjusted and adjusted curves are identical. Please see Section 4.3.5 and Appendix D.2.1 for further details on the adjustment method. Also, for more information on the formulation of conditional survival curves, please refer to Appendix D.2.2.

*Survey* (ACS) and *National Vital Statistics Reports* (NVSR) to estimate key mathematical parameters in the model. Through this empirical mapping, we (a) determine the years of life lost on the part of Black men due to excess incarceration versus excess mortality and (b) use the model to predict the number of years by which White men would adjust their schooling if they were faced with the same incarceration and mortality levels as Black men.

For an interest rate of two percent, the recent data show that 18-year-old (grade 12) Black males face a remaining lifetime — adjusted for incarceration and discounted to present value — of 30.54 years, compared to 33.45 years for age-18 White males. This gap represents an 8.69 percent lower marginal benefit of schooling for Black men. If we were to assume a value of 1 for the elasticity of schooling with respect to marginal benefit, as is supported by a range of literature, the Black-White gap in marginal benefit would suggest that Black men would invest 1.20 years ( $0.0869 \times 13.8$  years) less in schooling than White men. This approximation is remarkably close to the 1-year gap that we observe in the data.

In reality, the estimation cannot be as simple as assuming an elasticity value and performing this back-of-the-envelope calculation, because the elasticity itself depends on the marginal benefit of schooling. Therefore, we use our model of schooling to develop an estimation procedure for the elasticity, which then allows us to more rigorously estimate the schooling difference. Nevertheless, even with this procedure, we obtain elasticity estimates near unity, allowing us to explain most, if not all, of the Black/White schooling gap among males.

Though simple, the intuition underlying these findings has far-reaching implications. First, while many scholars and practitioners place a causal emphasis on education in reducing incarceration, the most significant causal relationship may in fact run in the opposite direction, with incarceration risk demotivating individuals from spending time in school. Second, the idea that an individual may invest less in education due to expectations of future incarceration embraces the fundamental notion in economics that everyone aims to do the best they can for themselves. When we recognize the agency exercised by individuals when faced with major incarceration disparities, we see the critical role of the disparities themselves in depressing educational achievement. Therefore, rather than “pushing” or “pulling” disadvantaged groups to obtain more schooling, more effective policies might focus on elimi-

nating such underlying social disparities and then simply allowing members of these groups to choose, for themselves, to invest more time in school.

The remainder of this article is structured as follows. In Section 4.2, we review the literature relevant to educational investment and incarceration. In Section 4.3, we develop a mathematical model of education that incorporates mortality and incarceration risk into the marginal benefit of schooling, and we describe the data that we use to estimate the model parameters. Then, in the Section 4.4, we present the model’s predictions of the schooling gap between Black and White men, along with the results of several robustness checks. Finally, Section 4.5 discusses the significance of the findings.

## 4.2 Literature Review

This study primarily relates to two key strands of literature. First, a body of work highlights the influence of incentives on educational investment. For example, with regard to the role of pecuniary incentives, [Dynarski \(2003\)](#) uses the elimination of the Social Security Student Benefit Program in the early 1980s to show that the availability of financial aid significantly impacts schooling outcomes. More directly related to this article, many studies analyze the responsiveness of education to changes in mortality. By shortening the time horizon over which the returns to education can be realized, mortality risk is thought to reduce the marginal benefit of schooling. Providing evidence for such a pathway, [Oster, Shoulson, and Dorsey \(2013\)](#) show that early-life confirmation of Huntington Disease, which significantly curtails life expectancy, leads to major reductions in human capital investment. In addition, throughout fifteen countries in sub-Saharan Africa, [Fortson \(2011\)](#) uses variation in HIV/AIDS prevalence over time and geography to demonstrate that individuals living in regions with higher HIV/AIDS burden experience significant declines in schooling outcomes. Several other studies obtain similar results.<sup>4</sup> [Bleakley \(2018\)](#) provides a comprehensive review of this so-called “horizon pathway” literature, demonstrating that all of the related studies arrive at an elasticity of schooling with respect to lifetime that is near or below unity.

Given that incarceration acts in a similar fashion as mortality by removing an individ-

---

<sup>4</sup>For example, see [Jayachandran and Lleras-Muney \(2009\)](#), [Baranov and Kohler \(2018\)](#), [Hansen \(2013\)](#), and [Stoler and Meltzer \(2013\)](#).



ual from the labor force, one might imagine that the risk of imprisonment also exhibits effects on education through the horizon mechanism. However, while a substantial body of literature — based on varying degrees of rigor — is devoted to the relationship between human capital, crime, and incarceration, no research addresses such a mechanism of action (to our knowledge). Rather, in contrast to the findings of this study, many scholars and practitioners place a causal emphasis on education (of some sort) in reducing incarceration. For instance, some stress the importance of schooling — and hence the elimination of the education gap — in preventing a path into criminal activity ([Lochner 2010](#); [Moretti 2005](#); [Lochner and Moretti 2004](#); [Lochner 2004](#); [Arum and Beattie 1999](#)), while others argue that there is a “school-to-prison” pipeline in which dysfunctional schools fail to provide such a path out of criminal activity ([Wilson 2014](#); [Raible and Irizarry 2010](#); [Hirschfield 2008](#)). In reality, the role of incarceration expectations in reducing educational attainment — which we highlight in this study — need not be incompatible with protective effects of education against incarceration. As we show below, our model can account for reverse causality from the education-to-incarceration pathway.

## 4.3 Methods

### 4.3.1 Conceptual Framework

How does an individual choose his education level? For each additional year invested in school, he should receive a higher wage later in life. Thus, an extra year of education increases income over the lifetime by some amount — that is, the marginal benefit of each year of school. But each additional year of school also comes at a cost: a direct cost from, say, tuition, and an opportunity cost from an individual foregoing wages that he would have otherwise earned in the labor force. Together, these costs make up the marginal cost of each additional year of school.

An individual will only choose to continue in school for an additional year if the marginal benefit of doing so outweighs the marginal cost. So long as the gap between the marginal benefit and marginal cost of schooling curves diminishes as schooling increases, so that

the two curves ultimately intersect, the individual will settle on an optimal educational investment. Specifically, he chooses his optimal years of education at the level of schooling for which the marginal benefit equals the marginal cost, stopping his education when the costs of continuing in school begin to outweigh the benefits.

Note that the marginal benefit of schooling is based on the future — via the accumulation of wages over the course of one’s lifetime — and depends essentially on available time in the labor force. With higher mortality and incarceration rates, Black men have a lower marginal benefit of schooling relative to White men, because their wage gains from additional education accrue to them over a shorter effective lifetime. Yet the marginal cost, which depends on costs incurred in the present, is not affected by this mechanism in the future.

We use the foregoing framework to compare educational choices in scenarios with Black male versus White male marginal benefits. This analysis yields an estimate of the primary quantity of interest in this study: the number of years by which Black men reduce their education due to excess incarceration and mortality. This quantity tells us the education gap between Black and White men that is attributable to disparities in effective lifetime.

### 4.3.2 Mathematical Model

Consider a representative agent born at  $t = 0$ . Suppose that the agent’s decision problem is to choose the number of years to invest in education that will maximize total income over the course of his lifetime (discounted to present value). To construct a measure of lifetime income, imagine that the agent’s wage is a function of education level. For an investment in  $s$  years of schooling (since birth), this wage,  $f(s)$ , accrues to the agent in each period of his life following period  $s$  — so long as the agent is alive and not incarcerated, and therefore in the labor force. To account for mortality and incarceration risks, we denote the probability of survival to period  $t$  as  $\ell(t)$  and the probability of incarceration in period  $t$  as  $m(t)$ . The term  $\ell(t)(1 - m(t))$  then gives survival adjusted for incarceration. Conditional on surviving to age  $s_0$ , survival adjusted for incarceration is given by  $\frac{\ell(t)(1-m(t))}{\ell(s_0)}$ .<sup>5</sup> In addition, each year

---

<sup>5</sup>In conditioning only on being alive at  $s_0$ , the model assumes that individuals who are incarcerated can still continue schooling while in prison. This assumption is consistent with options available to prisoners, including the ability to take the General Educational Development test and to pursue tertiary studies. Please see Appendix D.2.2 for further discussion of the implications of the conditional survival curve specification.

spent in school leading up to period  $s$  is associated with a direct cost — for instance, tuition — which we refer to as  $c(t)$ .

The lifetime income quantity is equal to the aggregation of all lifetime wages less the aggregation of the direct costs of school investment. Viewed from age  $s_0$ , it can be written as follows, using the foregoing notation:

$$Y \equiv \int_s^\infty f(s)e^{-(r-g)(t-s_0)} \frac{\ell(t)(1-m(t))}{\ell(s_0)} dt - \int_0^s c(t)e^{-(r-g)(t-s_0)} \frac{\ell(t)(1-m(t))}{\ell(s_0)} dt. \quad (4.1)$$

Note that  $r - g$  denotes the net interest rate, which serves to discount income and costs in each period of life to present value.

The agent seeks to maximize lifetime income by discontinuing schooling at some optimal number of years  $s^*$ , when the marginal cost of continuing for an additional year would outweigh the benefit of doing so. Mathematically, this idea corresponds to first-order differentiation of the lifetime income expression, which yields the following condition for the equalization of marginal benefit ( $MB$ ) and marginal cost ( $MC$ ) at the optimum:

$$MB = f'(s) \int_s^\infty e^{-(r-g)(t-s_0)} \frac{\ell(t)(1-m(t))}{\ell(s_0)} dt = f(s) + c(s) = MC. \quad (4.2)$$

In words, the marginal benefit of an additional year of school equals the extra lifetime income earned from the associated wage increase,  $f'(s)$ . Specifically, the following factor, which we call  $\Delta$ , represents incarceration-adjusted life years (discounted to present value) and inflates  $f'(s)$  to account for its aggregation over the lifetime:

$$\Delta(s) \equiv \int_s^\infty e^{-(r-g)(t-s_0)} \frac{\ell(t)(1-m(t))}{\ell(s_0)} dt. \quad (4.3)$$

The marginal cost of an additional year of school equals the direct cost of school for that year,  $c(s)$ , plus the opportunity cost of foregoing a year of wages in the labor force,  $f(s)$ . On the whole, the model is simple enough to be tractable but not too simple as to lose the core features of the agent's decision problem.

### 4.3.3 Estimation Procedure for Black-White Schooling Difference

Let a representative Black male agent choose an optimal  $s_b$  years of school and a representative White male agent choose an optimal  $s_w$  years. Because the White agent has a greater marginal benefit of schooling,  $s_w > s_b$ . We seek to estimate the quantity  $s_b - s_w$ , which represents the behavioral effect — due to re-optimization — of raising an individual’s mortality and incarceration rates from the White to the Black male levels.

Altering the marginal benefit of an additional year of schooling, mortality and incarceration enter the model by affecting the aforementioned  $\Delta$  term. Therefore, we use the model to predict  $s_b - s_w$ , the change in schooling due to a change in  $\Delta$  from the White male level to the Black male level. This quantity is given by the product of three terms: (1) the proportional difference in  $\Delta$  between Black and White men, (2) the initial schooling level of interest  $s_w$ , and (3) the elasticity of schooling with respect to  $\Delta$ , denoted as  $\varepsilon_{s_w}$ , which gives the proportional change in schooling at  $s_w$  due to a 0.01 proportional change in  $\Delta$ :

$$s_b - s_w = \left( \frac{\Delta_b - \Delta_w}{\Delta_w} \right) \varepsilon_{s_w} s_w. \quad (4.4)$$

The  $s_b - s_w$  prediction indicates the extent to which the Black-White male education gap is explained by incarceration and mortality disparities.

To estimate  $s_b - s_w$ , we recast equation (4.4) over the population of White males who have recently completed their education:

$$\widehat{s_b - s_w} = \sum_s p_s \cdot \max \left\{ \left( \frac{\Delta_b - \Delta_w}{\Delta_w} \right) \tilde{\varepsilon}_s s; \underline{s} - s \right\}, \quad (4.5)$$

where  $p_s$  is a weight given by the proportion of individuals with  $s$  years of education within the sample of non-Hispanic White men age 30-39,  $\tilde{\varepsilon}_s$  is an estimate of the elasticity at  $s$ , and  $\underline{s}$  denotes a lower bound on the schooling years to which the agent would adjust. This specification has three advantages. First, the resulting estimate represents the average population difference in schooling between Black and White men. Second, the weighted average helps to adjust for the non-linearity of the elasticity. Third, some estimates from the first argument of the maximum function are unrealistically large in magnitude, especially

given legal mandates for individuals to remain in school at younger ages. The  $\underline{s} - s$  argument accounts for these cases by bounding the adjustment. As shown below, we examine the results across  $\underline{s} \in \{6, 7, 8\}$ .

#### 4.3.4 Estimation Procedure for Elasticity

To calculate  $\tilde{\varepsilon}_s$ , we begin with the definition of the elasticity of schooling with respect to effective lifetime:

$$\varepsilon_s = \frac{\partial s}{\partial \Delta} \cdot \frac{\Delta}{s}. \quad (4.6)$$

Full differentiation of the lifetime income equation first-order condition ( $Y_s = 0$ ) yields

$$\frac{\partial s}{\partial \Delta} = \frac{f'}{-2f' \cdot \frac{\partial \Delta}{\partial s} - f \cdot \frac{\partial^2 \Delta}{\partial s^2} + c' - f'' \Delta}. \quad (4.7)$$

Thus, after dividing both the numerator and denominator of  $\frac{\partial s}{\partial \Delta}$  by  $f$  and combining the resulting expression with equation (4.6), we can rewrite the elasticity as

$$\varepsilon_s = \frac{f'/f}{-2 \cdot \frac{f'}{f} \cdot \frac{\partial \Delta}{\partial s} - \frac{\partial^2 \Delta}{\partial s^2} + \frac{c'}{f} - \frac{f''}{f} \cdot \Delta} \cdot \frac{\Delta}{s}. \quad (4.8)$$

As discussed in Section 4.2, a recent body of literature has estimated the elasticity of schooling with respect to  $\Delta$  to be about 1. We refer to this elasticity as  $\varepsilon_w$ , based on the assumption that it represents the standard elasticity of White men. Described as follows, we use  $\varepsilon_w \approx 1$  (along with  $\varepsilon_w \approx \frac{1}{2}$  and  $\varepsilon_w \approx 2$  to test for sensitivity) in a calibration procedure that determines the elasticity for White men under the hypothetical scenario in which they experience the same mortality and incarceration risks as Black men.

We begin with the observation that the quantity  $\frac{c'}{f} - \frac{f''}{f} \cdot \Delta$  in equation (4.8) is unknown under general conditions. To solve for this unknown quantity and to calibrate the elasticity, we use  $\varepsilon_w \approx 1$  in place of  $\varepsilon_s$ , along with estimates from the data of  $\Delta$ ,  $\frac{\partial \Delta}{\partial s}$ , and  $\frac{\partial^2 \Delta}{\partial s^2}$  for White men, all evaluated at some initial schooling level and age  $s_0$ .<sup>6</sup> The equation for  $\Delta$  is

<sup>6</sup>In the model, age and education level are one and the same, since schooling is measured in terms of years since birth. When we map the model to data, we assume that  $age = s + 6$ . Thus, in our estimation procedure, since survival curves are defined by age, the conditioning age  $s_0$  that is used for the survival curve normalization is six years greater than the grade level that is used for the lower bound of the  $\Delta$  integral.

shown above, while equations for  $\frac{\partial \Delta}{\partial s}$  and  $\frac{\partial^2 \Delta}{\partial s^2}$  are shown as follows. All three quantities can be estimated with survival and incarceration data.

$$\frac{\partial \Delta}{\partial s} = -e^{-(r-g)(s-s_0)} \cdot \frac{\ell(s)(1-m(s))}{\ell(s_0)} - \int_s^\infty e^{-(r-g)(t-s_0)} \cdot \frac{\ell(t)}{\ell(s_0)} \cdot \frac{\partial m(t)}{\partial s} dt \quad (4.9)$$

$$\left. \frac{\partial \Delta}{\partial s} \right|_{s_0} = -1 + m(s_0) \quad (4.10)$$

$$\frac{\partial^2 \Delta}{\partial s^2} = (r-g)e^{-(r-g)(s-s_0)} \cdot \frac{\ell(s)(1-m(s))}{\ell(s_0)} - e^{-(r-g)(s-s_0)} \cdot \left[ \frac{\partial \ell(s)}{\partial s} \cdot \frac{(1-m(s))}{\ell(s_0)} \right] \quad (4.11)$$

$$\begin{aligned} \left. \frac{\partial^2 \Delta}{\partial s^2} \right|_{s_0} &= (r-g)(1-m(s_0)) - \frac{\partial \ell(s_0)}{\partial s} \cdot \frac{(1-m(s_0))}{\ell(s_0)} \\ &= (r-g)(1-m(s_0)) \end{aligned} \quad (4.12)$$

Next, for  $\frac{f'}{f}$  — equal to Mincer's  $\beta$  coefficient, measuring the percentage gain in wages from an additional year of school — we employ estimates of about 0.1, which have been found in countless studies.

Using the foregoing estimates, we employ two approaches to calibrating the elasticity. In the first, more general case, which we refer to as Elasticity Method 1, we insert the White-male  $\Delta$ ,  $\frac{\partial \Delta}{\partial s}$ , and  $\frac{\partial^2 \Delta}{\partial s^2}$  estimates into equation (4.8) for the elasticity, yielding an estimate for the unknown quantity  $\frac{c'}{f} - \frac{f''}{f} \cdot \Delta$ :

$$\frac{c'}{f} - \frac{f''}{f} \cdot \Delta \approx \frac{\beta \Delta_w}{s} + 2\beta \cdot \frac{\partial \Delta_w}{\partial s} + \frac{\partial^2 \Delta_w}{\partial s^2}, \quad (4.13)$$

where the  $w$  subscripts on the  $\Delta$  terms indicate that they correspond to the estimates for White men. In the second approach, Elasticity Method 2, we follow the finding from the literature of log-linear  $f(s)$  (Card 1999), which implies that  $\frac{f''}{f} = \beta^2$ . This relationship allows us to isolate  $\frac{c'}{f}$  as the unknown quantity:

$$\frac{c'}{f} \approx \frac{\beta \Delta_w}{s} + 2\beta \cdot \frac{\partial \Delta_w}{\partial s} + \frac{\partial^2 \Delta_w}{\partial s^2} + \frac{f''}{f} \cdot \Delta_w. \quad (4.14)$$

In both approaches, we then input into the elasticity expression estimates of the respective unknown quantity and estimates of  $\Delta$ ,  $\frac{\partial\Delta}{\partial s}$ , and  $\frac{\partial^2\Delta}{\partial s^2}$  for Black men. Putting everything together, we have the following for the elasticity estimate under Elasticity Method 1:

$$\tilde{\varepsilon}_s = \frac{\beta}{\frac{\beta\Delta_w}{s} + 2\beta\left(\frac{\partial\Delta_w}{\partial s} - \frac{\partial\Delta_b}{\partial s}\right) + \frac{\partial^2\Delta_w}{\partial s^2} - \frac{\partial^2\Delta_b}{\partial s^2}} \cdot \frac{\Delta_b}{s}. \quad (4.15)$$

Using Elasticity Method 2, we have

$$\tilde{\varepsilon}_s = \frac{\beta}{\frac{\beta\Delta_w}{s} + \beta^2(\Delta_w - \Delta_b) + 2\beta\left(\frac{\partial\Delta_w}{\partial s} - \frac{\partial\Delta_b}{\partial s}\right) + \frac{\partial^2\Delta_w}{\partial s^2} - \frac{\partial^2\Delta_b}{\partial s^2}} \cdot \frac{\Delta_b}{s}. \quad (4.16)$$

These elasticity estimates characterize how White men would adjust their years in school if they faced the same mortality and incarceration constraints as Black men. Using the estimated elasticity and the White and Black values for  $\Delta$ , we are then able to estimate equation (4.5).

### 4.3.5 Data

We obtain mortality (by age, sex, and race) data from the NVSR to construct the survival curves that enter into the analysis. From the ACS, we use data on age, sex, race, education, and institutionalization. For the teenage and younger adult years, the ACS's aggregated category of institutionalization almost entirely reflects incarceration. For the older adult years, however, the institutionalization measure increases significantly, primarily due to elderly-related institutionalization. Therefore, we institute an age cutoff of 60 for the survival probability adjustment. For ages strictly above the cutoff, we assume that the incarceration rate  $m$  becomes zero, and we no longer perform the adjustments to the standard survival curve. The choice of cutoff is based on the analysis of ACS data from 1980, the last year in which the ACS collected disaggregated data on institutionalization that were incarceration-specific. Finally, note that because all of our estimates are based on lifespan from some teenage or young adult age  $s_0$ , we do not need to address the presence of any non-incarceration institutionalization during the childhood years.<sup>7</sup>

---

<sup>7</sup>See Appendix D.2.1 for additional discussion of the survival probability adjustment method.

## 4.4 Results

We first examine differences in  $\Delta(s)$  — remaining lifetime, discounted to present value — between Black and White men. Table 4.1 displays  $\Delta(s)$  figures, both unadjusted and adjusted for incarceration, along with the proportional difference between the adjusted figures for males of each race.<sup>8</sup> Our tabulations show a significant disparity. Specifically, for a two-percent time discount rate, Black males face an incarceration-adjusted remaining lifetime of 30.54 years, compared to 33.45 years for age-18 White males. Because effective life years determine the marginal benefit of schooling, this disparity means that the returns to schooling for Black men are 8.69 percent lower than for White men.

TABLE 4.1: REMAINING LIFETIME AT AGE 18 FOR BLACK AND WHITE MEN, WITH AND WITHOUT INCARCERATION ADJUSTMENT

	(1)	(2)	(3)	(4)	(5)
$r - g$	$\Delta_b^u$	$\Delta_w^u$	$\Delta_b$	$\Delta_w$	$\left(\frac{\Delta_b - \Delta_w}{\Delta_w}\right)$
-1%	76.25	82.66	72.73	81.86	-0.112
0%	55.40	59.15	52.56	58.53	-0.102
2%	32.49	33.85	30.54	33.45	-0.0869
5%	18.10	18.46	16.88	18.22	-0.0734

*Notes:* This table displays estimates of the remaining number of life years from the age of 18, across net interest rate  $r - g$ , for Black and White men. Columns (1) and (2), where the  $\Delta$  figure includes the  $u$  superscript, provide remaining lifetime figures based on mortality rates alone — that is, unadjusted for incarceration. Columns (3) and (4) display  $\Delta$  figures adjusted for incarceration, based on equation (4.3). Column (5) shows the proportional difference in the incarceration-adjusted  $\Delta$  figures between Black and White men.

Our next set of results shows that this inequality in marginal benefit explains a substantial portion of the education gap between Black and White men. Table 4.2 displays our main results for the model’s predicted schooling difference, across net interest rate and choice of

<sup>8</sup>The unadjusted version of  $\Delta(s)$  is equivalent to that provided in equation (4.3) but without the  $(1 - m(t))$  term.



White-male elasticity for the estimation procedure. For a net interest rate of two percent and a choice of  $\varepsilon_w = 1$ , the model predicts that Black men will obtain 1.27 years less education than White men, slightly over-explaining the one-year gap between the two groups. This estimate is very close to the 1.20-year result from the back-of-the-envelope calculation presented in Section 4.1.<sup>9</sup> Naturally, as the interest rate increases, the explained difference decreases, since greater time discounting leads to a smaller spread in effective lifetime between Black and White men. As the assumed elasticity for White men increases, the explained schooling difference increases, since the greater elasticity signifies larger adjustment in response to changes in marginal benefit. Figure 4.3 provides a graphical representation of the main results, showing the predicted change in schooling, for each initializing elasticity, across net interest rate.

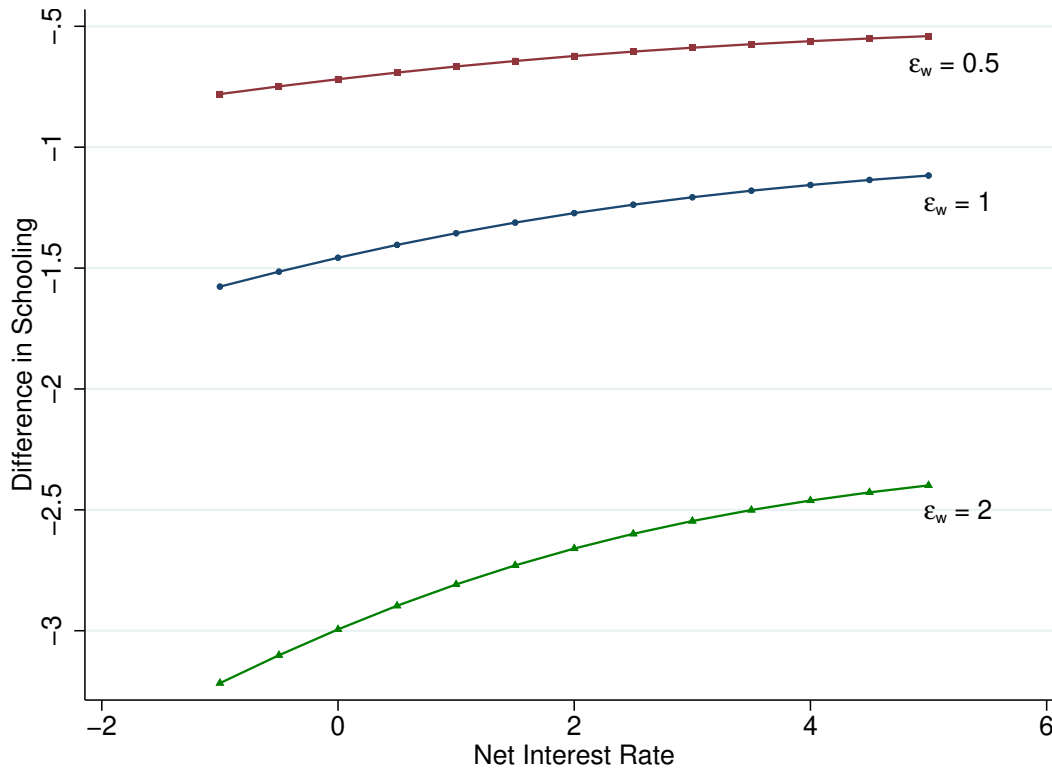
TABLE 4.2: PREDICTED CHANGE IN YEARS OF SCHOOLING ACROSS NET INTEREST RATE AND ASSUMED WHITE-MALE ELASTICITY

	(1)	(2)	(3)
$r - g$	$\varepsilon_w = \frac{1}{2}$	$\varepsilon_w = 1$	$\varepsilon_w = 2$
-1%	-0.780	-1.577	-3.217
0%	-0.719	-1.457	-2.994
2%	-0.623	-1.273	-2.660
5%	-0.541	-1.117	-2.399

*Notes:* This table displays estimates of the schooling difference between Black and White Men across two parameters of the model: net interest rate  $r - g$  and assumed White-male elasticity  $\varepsilon_w$ . The estimated schooling difference is computed as the number of years by which White men would reduce their educational investment if they experienced the same mortality and incarceration rates as Black men. Specifically, the figures in the table correspond to estimates of  $s_b - s_w$  from equation (4.5). The base specification of the model uses  $r - g = 0.02$  and  $\varepsilon_w = 1$ .

<sup>9</sup>If we redo the back-of-the-envelope calculation using the elasticity estimate of  $\tilde{\varepsilon}_s = 1.02$  that we obtain for a starting point of 14 years of school (as the full year nearest to the 13.8-year average educational attainment of White men), the estimated schooling gap would be 1.24 years.

FIGURE 4.3: ESTIMATED SCHOOLING DIFFERENCES FOR VARIOUS ELASTICITIES



*Notes:* This figure graphs the estimates of the schooling difference between Black and White Men, across (a) different choices for the assumed value of the White-male elasticity  $\epsilon_w$  and (b) net interest rate. The curves consist of points corresponding to a domain of thirteen net interest rates, ranging from -1 to 5 percent by increments of 0.5 percentage points. The estimated schooling difference is computed as the number of years by which White men would reduce their educational investment if they experienced the same mortality and incarceration rates as Black men. Specifically, the curves correspond to estimates of  $s_b - s_w$  from equation (4.5). The base specification of the model uses a net interest rate of 2 percent and  $\epsilon_w = 1$ .

#### 4.4.1 Robustness Checks

We next carry out several checks to explore the robustness of the findings to various properties of the model and our estimation methods. These checks include the possible role of education in protecting against incarceration, as well as changes to the data specifications and estimation procedure. The checks do not substantively alter the core results.

##### 4.4.1.1 Incarceration-Education Circularity

While our model is predicated on the incarceration-to-education pathway, what happens if education protects against incarceration? The presence of such circularity between education and incarceration could confound our estimates of the incarceration-on-education effects. If

an agent's educational choice reduces incarceration risk, the model above would produce overestimates of the portion of the Black-White educational gap explained by incarceration, since in reality the protective effects of education would drive agents to choose higher levels of schooling, thereby counteracting the effect of the incarceration expectations pathway.

We therefore extend the model to account for this possibility. Specifically, we consider incarceration  $m$  as a function of not only age  $t$  but schooling  $s$ . When an individual determines his optimal schooling investment, then, his choice will also account for the *gains* in  $\Delta$  that he may experience as a result of the protective effect of education against incarceration. The  $m(t, s)$  version of the model leads to the following equations for  $\Delta(s)$ ,  $\frac{\partial \Delta}{\partial s}$ , and  $\frac{\partial^2 \Delta}{\partial s^2}$ .

$$\Delta(s) \equiv \int_s^\infty e^{-(r-g)(t-s_0)} \frac{\ell(t)(1-m(t,s))}{\ell(s_0)} dt \quad (4.17)$$

$$\frac{\partial \Delta}{\partial s} = -e^{-(r-g)(s-s_0)} \cdot \frac{\ell(s)(1-m(s,s))}{\ell(s_0)} - \int_s^\infty e^{-(r-g)(t-s_0)} \cdot \frac{\ell(t)}{\ell(s_0)} \cdot \frac{\partial m(t,s)}{\partial s} dt \quad (4.18)$$

$$\left. \frac{\partial \Delta}{\partial s} \right|_{s_0} = -1 - \int_{s_0}^\infty e^{-(r-g)(t-s_0)} \cdot \frac{\ell(t)}{\ell(s_0)} \cdot \frac{\partial m(t,s_0)}{\partial s} dt \quad (4.19)$$

$$\begin{aligned} \frac{\partial^2 \Delta}{\partial s^2} &= (r-g)e^{-(r-g)(s-s_0)} \cdot \frac{\ell(s)(1-m(s,s))}{\ell(s_0)} \\ &\quad - e^{-(r-g)(s-s_0)} \cdot \left[ \frac{\partial \ell(s)}{\partial s} \cdot \frac{(1-m(s,s))}{\ell(s_0)} - \frac{\partial m(s,s)}{\partial s} \cdot \frac{\ell(s)}{\ell(s_0)} \right] \\ &\quad + e^{-(r-g)(s-s_0)} \cdot \frac{\ell(s)}{\ell(s_0)} \cdot \frac{\partial m(s,s)}{\partial s} \\ &\quad - \int_{s_0}^\infty e^{-(r-g)(t-s_0)} \cdot \frac{\ell(t)}{\ell(s_0)} \cdot \frac{\partial^2 m(t,s)}{\partial s^2} dt \end{aligned} \quad (4.20)$$

$$\begin{aligned} \frac{\partial^2 \Delta}{\partial s^2} \Big|_{s_0} &= r - g - \frac{\partial \ell(s_0)}{\partial s} \cdot \frac{(1 - m(s_0, s_0))}{\ell(s_0)} + 2 \cdot \frac{\partial m(s_0, s_0)}{\partial s} \\ &\quad - \int_{s_0}^{\infty} e^{-(r-g)(t-s_0)} \cdot \frac{\ell(t)}{\ell(s_0)} \cdot \frac{\partial^2 m(t, s_0)}{\partial s^2} dt \end{aligned} \quad (4.21)$$

Equation (4.18) helps to illustrate the potential protective effect of schooling. If schooling reduces incarceration, then  $\frac{\partial m}{\partial s} < 0$ . The integration of  $\frac{\partial m}{\partial s}$  terms will then have the effect of increasing  $\frac{\partial \Delta}{\partial s}$ , which corresponds to an increase in  $\Delta$  through the effect of school choice on incarceration.

Similar to the estimation procedure above, all of the foregoing quantities can be estimated with survival, institutionalization, education, and age data. For the first and second derivatives of  $m(t, s)$ , we compute estimates using a flexible, 4<sup>th</sup>-degree polynomial regression of incarceration on age and education.<sup>10</sup> To produce estimates of the predicted educational gap incorporating the possible effect of education on incarceration risk, we employ a weighted average between  $m(t)$  and  $m(t, s)$  in the computation of  $\Delta(s)$ . The resulting form of  $\Delta(s)$ , which we subscript by the weight  $\alpha$ , is equivalent to a weighted average of  $\Delta$  computed using  $m(t)$  and  $\Delta$  computed using  $m(t, s)$ :

$$\begin{aligned} \Delta_{\alpha}(s) &= \int_s^{\infty} e^{-(r-g)(t-s)} \cdot \frac{\ell(t) [1 - ((1 - \alpha) \cdot m(t) + \alpha \cdot m(t, s))]}{\ell(s_0)} dt \\ &= \alpha \cdot \Delta_{m(t,s)} + (1 - \alpha) \cdot \Delta_{m(t)}. \end{aligned} \quad (4.22)$$

Note that  $\alpha = 0$  corresponds to the default model in which schooling has no effect on incarceration. When  $\alpha = 1$ , the setup places full weight on the gradient between education and incarceration observed in the data; in other words,  $\alpha = 1$  represents an upper bound on the possible effect of education on incarceration.

Using a range of choices for  $\alpha \in [0, 1]$ , Table 4.3 shows the results for this extension of the model that incorporates a dependency of incarceration on education. Each of the  $\alpha = 0$  columns corresponds to the base case shown in Table 4.2. As  $\alpha$  increases from 0 to 1, placing increasing weight on  $m(t, s)$ , the magnitude of the predicted gap in education tends to decrease, in accordance with the notion that education protects against incarceration.

---

<sup>10</sup>See Appendix D.3.2 for further detail.

However, the relationship is not strictly monotonic. Moreover, the change in the schooling gap estimate from  $\alpha = 0$  to  $\alpha = 1$  is relatively small. For example, for a 2% interest rate and  $\varepsilon_w = 1$ , the 1.27-year base case estimate decreases by about 0.44 years to 0.83 years in the estimate based on the full weight of the observed education-on-incarceration gradient. This adjusted estimate still explains almost all of the observed Black-White male schooling gap. In addition, estimates based on weights of  $\alpha = \frac{1}{3}$  and  $\alpha = \frac{2}{3}$ , which may provide more realistic representations of the protective effects of education, come remarkably close to the actual 1-year education difference. All in all, the marginal effect of the education-to-incarceration pathway on the overall predicted gap in schooling accords with intuition: the difference in  $\Delta$  between Black and White men produces a major first-order effect on schooling choice, whereas the influence of schooling choice on  $\Delta$  via its reduction of incarceration risk amounts to a second-order effect on the educational gap.

#### 4.4.1.2 Data and Estimation Procedure Specifications

We also check for robustness to several specification choices made in the data and estimation procedure. First, in the base case, we use  $\underline{s} = 7$  in equation (4.5). As shown in Table 4.4, we test for sensitivity around this value by also using  $\underline{s} = 6$  and  $\underline{s} = 8$ . The estimated educational difference barely budges, implying that bounding of the schooling adjustment by  $\underline{s} - s$  has little effect on the estimates.

Second, in our base case, we code the General Educational Development (GED) test as grade 12, while it may instead be coded as grade 11. Additionally, we employ equation (4.13) to calibrate the elasticity, whereas we may instead use equation (4.14) on the assumption that  $\frac{f''}{f} = \beta^2$ . Table 4.5 displays the results when we vary these different assumptions, under the range of choices for the education-on-incarceration gradient weight  $\alpha$ . The grade level choice for GED coding has a minimal effect on the estimates, while the  $\frac{f''}{f}$  assumption reduces the explained education gap by around 0.2 years in most of the cases.

TABLE 4.3: PREDICTED CHANGE IN YEARS OF SCHOOLING ACROSS NET INTEREST RATE, ASSUMED WHITE-MALE ELASTICITY, AND EDUCATION-ON-INCARCERATION GRADIENT WEIGHT

	$\varepsilon_w = \frac{1}{2}$			$\varepsilon_w = 1$			$\varepsilon_w = 2$					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
$r - g$	$\alpha = 0$	$\alpha = \frac{1}{3}$	$\alpha = \frac{2}{3}$	$\alpha = 1$	$\alpha = 0$	$\alpha = \frac{1}{3}$	$\alpha = \frac{2}{3}$	$\alpha = 1$	$\alpha = 0$	$\alpha = \frac{1}{3}$	$\alpha = \frac{2}{3}$	$\alpha = 1$
-1%	-0.780	-0.683	-0.628	-0.598	-1.577	-1.301	-1.210	-1.184	-3.217	-2.456	-2.357	-2.459
0%	-0.719	-0.612	-0.556	-0.529	-1.457	-1.161	-1.078	-1.058	-2.994	-2.188	-2.103	-2.143
2%	-0.623	-0.499	-0.443	-0.422	-1.273	-0.940	-0.860	-0.830	-2.660	-1.756	-1.567	-1.595
5%	-0.541	-0.396	-0.338	-0.319	-1.117	-0.734	-0.642	-0.523	-2.399	-1.326	-1.036	-0.871

*Notes:* This table displays estimates of the schooling difference between Black and White Men across three parameters of the model: net interest rate  $r - g$ , assumed White-Male elasticity  $\varepsilon_w$ , and education-on-incarceration gradient weight  $\alpha$ . The table expands upon Table 4.2 with the inclusion of the  $\alpha$  parameter. The estimated schooling difference is computed as the number of years by which White men would reduce their educational investment if they experienced the same mortality and incarceration rates as Black men. Specifically, the figures in the table correspond to estimates of  $s_b - s_w$  from equation (4.5). The base specification of the model uses  $r - g = 0.02$ ,  $\varepsilon_w = 1$ , and  $\alpha = 0$ .

TABLE 4.4: PREDICTED CHANGE IN YEARS OF SCHOOLING ACROSS NET INTEREST RATE AND SCHOOLING LOWER BOUND

	(1)	(2)	(3)
$r - g$	$\underline{s} = 6$	$\underline{s} = 7$	$\underline{s} = 8$
-1%	-1.578	-1.577	-1.573
0%	-1.458	-1.457	-1.454
2%	-1.273	-1.273	-1.270
5%	-1.118	-1.117	-1.115

*Notes:* This table displays estimates of the schooling difference between Black and White Men across two parameters of the model: net interest rate  $r - g$  and schooling adjustment bound  $\underline{s}$ . The estimated schooling difference is computed as the number of years by which White men would reduce their educational investment if they experienced the same mortality and incarceration rates as Black men. Specifically, the figures in the table correspond to estimates of  $s_b - s_w$  from equation (4.5). The base specification of the model uses  $r - g = 0.02$  and  $\underline{s} = 7$ .

## 4.5 Discussion

In this article, we begin with the premise that the higher burden of incarceration and mortality experienced by Black men relative to White men depresses their marginal benefit of schooling by reducing their effective lifetimes, especially their available years in the labor market. We hypothesize that, as a result of the enhanced constraints, Black men invest less time in school. To test this hypothesis, we construct a model of the educational investment decision in the setting of incarceration and mortality constraints and use available data to estimate the model’s predicted gap in schooling between Black and White men.

Our estimates show that incarceration and mortality inequalities constitute a key, under-recognized explanation for the major education disparities between Black and White men, accounting for a substantial portion of the one-year racial gap. Specifically, in our preferred base specification of the schooling model, we estimate a schooling difference of 1.20 years, which slightly over-explains the education gap. This finding is robust to a range of model and data specification checks. At a 2% time-discount rate and with a value of 1 for the

TABLE 4.5: PREDICTED CHANGE IN YEARS OF SCHOOLING ACROSS ELASTICITY METHODS AND GED CODING

	GED = 12		GED = 11	
	(1)	(2)	(3)	(4)
$\alpha$	General	$\frac{f''}{f} = \beta^2$	General	$\frac{f''}{f} = \beta^2$
0	-1.273	-1.024	-1.268	-1.020
$\frac{1}{3}$	-0.940	-0.783	-0.920	-0.765
$\frac{2}{3}$	-0.860	-0.714	-0.860	-0.705
1	-0.830	-0.695	-0.819	-0.709

*Notes:* This table displays estimates of the schooling difference between Black and White Men across (a) elasticity estimation method, (b) coding of the GED, and (c) education-on-incarceration gradient weight  $\alpha$ . The estimated schooling difference is computed as the number of years by which White men would reduce their educational investment if they experienced the same mortality and incarceration rates as Black men. Specifically, the figures in the table correspond to estimates of  $s_b - s_w$  from equation (4.5). The “General” columns correspond to Elasticity Method 1, which is given by equation (4.13). The  $\frac{f''}{f} = \beta^2$  columns correspond to Elasticity Method 2, which is given by equation (4.14). All estimates in the table correspond to net interest rate  $r - g = 0.02$ . The base specification of the model uses Elasticity Method 1, GED = 12 years, and  $\alpha = 0$ .

White-male elasticity of schooling with respect to effective lifetime, our robustness checks show that the model’s minimum predicted gap is 0.70 years — still accounting for about 70% of the education disparity. Moreover, when we incorporate into the model the possible effect of education on incarceration risk, we obtain estimates that predict the observed one-year Black-White gap with remarkable precision.

These results have important implications for our understanding of human capital investment and for policies aimed at addressing racial disparities in education. Most of all, our findings cast significant doubt on the oft-expressed “school prevents incarceration” perspective, which suggests that policymakers should primarily focus on methods for keeping individuals in school. This perspective takes little account of the agency of these individuals and the environmental constraints affecting their choices. In fact, if a given individual has chosen an *optimal* level of schooling, then keeping him in school longer will shift him away from his optimum and *reduce* his welfare — a counter-intuitive but important reality. Thus,



while informative, the overall perspective of education as a protector against incarceration does not account for the full picture of human behavior and welfare, because it does not address the question of why individuals may choose low levels of education in the first place.

In contrast, the results of the choice-based approach in this study illuminate the need to address underlying constraints — higher mortality and incarceration risks of Black men — to resolve education disparities. We arrive at these results even while incorporating the protective effect of education into our model. Thus, altogether, it may be that the best and perhaps only way to raise the educational attainment of Black men is to address the social inequalities that underpin their choices — and then simply to allow them to choose more school.

## References

- Arum, Richard, and Irene R Beattie. 1999. "High School Experience and the Risk of Adult Incarceration." *Criminology* 37 (3): 515–540.
- Baranov, Victoria, and Hans-Peter Kohler. 2018. "The Impact of AIDS Treatment on Savings and Human Capital Investment in Malawi." *American Economic Journal: Applied Economics* 10 (1): 266–306.
- Bleakley, Hoyt. 2018. *Longevity, Education, and Income: How Large is the Triangle?* Tech. rep. National Bureau of Economic Research.
- Card, David. 1999. "The Causal Effect of Education on Earnings." In *Handbook of Labor Economics*, 3:1801–1863. Elsevier.
- Dynarski, Susan M. 2003. "Does Aid Matter? Measuring the Effect of Student Aid on College Attendance and Completion." *American Economic Review* 93 (1): 279–288.
- Fortson, Jane G. 2011. "Mortality Risk and Human Capital Investment: The Impact of HIV/AIDS in Sub-Saharan Africa." *The Review of Economics and Statistics* 93 (1): 1–15.
- Hansen, Casper Worm. 2013. "Life Expectancy and Human Capital: Evidence from the International Epidemiological Transition." *Journal of Health Economics* 32 (6): 1142–1152.
- Hirschfield, Paul J. 2008. "Preparing for Prison? The Criminalization of School Discipline in the USA." *Theoretical Criminology* 12 (1): 79–101.
- Jaeger, David A. 2003. "Estimating the Returns to Education Using the Newest Current Population Survey Education Questions." *Economics Letters* 78 (3): 385–394.
- . 1997. "Reconciling the Old and New Census Bureau Education Questions: Recommendations for Researchers." *Journal of Business & Economic Statistics* 15 (3): 300–309.
- Jayachandran, Seema, and Adriana Lleras-Muney. 2009. "Life Expectancy and Human Capital Investments: Evidence from Maternal Mortality Declines." *The Quarterly Journal of Economics* 124 (1): 349–397.
- Lochner, Lance. 2010. "Education and Crime." In *International Encyclopedia of Education*, 3rd ed., ed. by Penelope Peterson, Rob Tierney, Eva Baker, and Barry McGaw, 239–244. Amsterdam, Netherlands: Elsevier Science.
- . 2004. "Education, Work, and Crime: A Human Capital Approach." *International Economic Review* 45 (3): 811–843.
- Lochner, Lance, and Enrico Moretti. 2004. "The Effect of Education on Crime: Evidence from Prison Inmates, Arrests, and Self-Reports." *American Economic Review* 94 (1): 155–189.
- Moretti, Enrico. 2005. "Does Education Reduce Participation in Criminal Activities?" In *Symposium on "The Social Costs of Inadequate Education"* (Columbia University Teachers College).

- Oster, Emily, Ira Shoulson, and E Dorsey. 2013. "Limited Life Expectancy, Human Capital and Health Investments." *American Economic Review* 103 (5): 1977–2002.
- Raible, John, and Jason G Irizarry. 2010. "Redirecting the Teacher's Gaze: Teacher Education, Youth Surveillance and the School-to-Prison Pipeline." *Teaching and Teacher Education* 26 (5): 1196–1203.
- Stoler, Avraham, and David Meltzer. 2013. "Mortality and Morbidity Risks and Economic Behavior." *Health Economics* 22 (2): 132–143.
- Wilson, Harry. 2014. "Turning Off the School-to-Prison Pipeline." *Reclaiming Children and Youth* 23 (1): 49.

## CHAPTER 5

### Quality Production in the Development Sector

#### 5.1 Introduction

Quality is a major running theme throughout this dissertation. Chapter 2 addresses matters of quality most directly, focusing on how to improve governance incentives to foster the delivery of higher-quality health services. Some of the conclusions in Chapter 3 also land on matters of quality, suggesting that low-quality intervention delivery by community health workers may have explained the study's largely null results, and that effective management must be an emphasis moving forward for community health worker programs.

In addition to the quality-related content that is explicitly covered in the dissertation, Chapters 1, 2, and 3 all relate to projects in Uganda that have required a high level of management input over the course of implementation. Some critics of this work may question its generalizability, observing that the intensity of program implementation (particularly for Chapters 1 and 2) may not be easily replicated in other settings. While such an observation is not entirely wrong, this criticism misses the bigger picture — for the fundamental question we should be asking ourselves is the following: what barriers are stopping the programs from being easily replicated in other settings? Rather than writing off interventions as impractical, we should strategize to change the core environmental factors that render certain activities less feasible — in the same fashion as the work in Chapter 2 attempts to alter the political environment to promote better delivery of proven health services. We can think of such an endeavor as an effort to raise the effectiveness of interventions to their level of efficacy, rather than discarding efficacious interventions out of resignation that they will never be effective. Put simply, our response should not be to drop our standards, but instead to insist that those standards be met.

In fact, in almost all regards, there is nothing special about Progressive Health Partnership (PHP), the budding organization operating on a shoestring budget that carried out the project activities for Chapters 1, 2, and 3. Yet, based on anecdotal evidence, the organization implemented the activities with a level of quality that far exceeded that of other well-known entities with budgets many times the size of its own. If anything makes PHP special, perhaps it is the systems and incentives that the organization has established to shape its work, highlighting the fact that quality often is not about the level of available resources but the use of those resources. Nonetheless, PHP's very act of implementing its programs serves as proof of concept that high quality standards can be achieved. After all, few if any of PHP's practices are revolutionary; they are well within the reach of many different entities. Much of the time, however, the adherence to quality standards hinges on underlying incentives.

By studying these incentives, and by identifying the systems that produce quality, we can make important strides. Along these lines, in the very process of implementing this dissertation's projects and working in the development sector more generally, I have distilled a number of critical lessons about quality. These lessons have arisen from my fundamental observation of the substandard work that pervades the development industry. It is all too common to see large sums of money invested in hollow programs that create little to no impact or even produce unintended consequences. In other instances, development organizations demonstrate substantial skill in building visible structures — such as health facilities or schools — but do much less to establish appropriate systems to deliver quality services from those structures. Why do we not demand more? If manufacturers across many for-profit sectors can fabricate products with near-zero defect rates, why do we not demand similar standards from the development sector?

The complete answer to this question goes beyond the scope of this chapter and is related to why, at a macroscopic level, the development sector is not better structured to produce quality. In short, the explanation again boils down to incentives, including the entanglement of interests and power dynamics that distort production incentives away from the ultimate users of services. However, part of the answer, which I address herein, concerns the complex nature of quality itself. As an object, quality, which often includes intangible elements, can be difficult to pin down. So how can we think about quality at a fundamental level? What is

it, and how is it created? The following sections provide several key lessons touching on these questions. Specifically, Section 5.2 discusses difficulties in measuring and incentivizing quality, ultimately concluding that we need to shift our view of quality in the development sector toward the upstream incentives and institutions that create it and away from the downstream outcomes and products of aid programs. Section 5.3 then presents a simple mathematical model of production cycle quality — which can be applied to virtually any phase-based task — to demonstrate the important roles of quality assurance and quality control. Following, Section 5.4 combines the lessons of the two preceding sections to illustrate selection as a key strategy for achieving workforce quality, particularly in development organizations. Finally, Section 5.5, using a real-life application, builds on the foregoing observations by examining the role of institutions — or leadership — as another determinant of workforce quality. The section then draws the entire chapter to a close with the conclusion that governance is the institutionalization of leadership itself.

## 5.2 Quality as Properly Aligned Incentives

Quality is an intriguing and elusive object. Not only does quality tend to be multidimensional, but at least some of its dimensions tend to be unobservable to those with an interest in judging the quality of goods and services. In this regard, the development sector, which is rife with principal-agent relationships and information asymmetries at multiple levels, is especially complex from a quality standpoint. For instance, at a political level, voters cannot fully observe the quality of the politicians whom they contract to represent them. In turn, the politicians cannot fully observe the quality of the bureaucrats whom they contract to implement government policy in the service of voters. At the level of service provision, managers cannot fully observe the quality of the employees whom they contract to provide services. Likewise, service beneficiaries do not have sufficient information to adequately assess the quality of the service providers.

While this collection of different actors suggests that we can view quality from a variety of standpoints, this section adopts the perspective of a third-party observer who seeks to objectively assess the quality of entities in the development sector. The observer, as a neutral

party, seeks to maximize social welfare, and we can therefore conceive of this viewpoint as representing the preferences of society. In employing this perspective, I analyze various dynamics of the development sector in relation to quality.

Overall, the multidimensional and often unobservable nature of quality make the measurement of quality — in conjunction with attempts to motivate quality — a complicated undertaking. Even when the dimensions of quality are completely observable, it may be costly to regularly measure all of them. But given the unobservable components of quality, it can be challenging, if not impossible, to develop a valid measure of the object. To illustrate the core dilemma, consider the task of a social worker whose job is to visit malnourished children and their families. The overall quality of the worker's performance might be based on multiple factors, such as the number of visits made to households as well as the effectiveness of those visits. The former is easily measured, while the latter is much more difficult to measure, since it may include intangible components such as the quality of rapport-building and education delivery. Because of these unobservable (and noncontractible) dimensions of quality, it is inherently difficult to incentivize the worker's performance. If the worker is rewarded based on the number of households covered, she will naturally allocate more effort toward achieving high volume, trading off with the quality of the home visits themselves. Alternatively, if the worker's firm does not have sound monitoring mechanisms in place, the worker may even report a fabricated number of home visits to earn a larger reward. Similar concerns to these observations have been expressed regarding teacher incentives based on standardized test scores, which may operate at the expense of more intangible, higher-order learning and developmental outcomes, such as creative thinking and socialization.

[Holmstrom and Milgrom \(1991\)](#) have formalized many such observations in the context of principal-agent problems where the agent's task is multidimensional. When tasks consist of multiple components, incentives serve not only to motivate hard work but also to allocate effort *among* the different dimensions. Building on this principle, the authors show that optimal contracts in these settings may offer a fixed wage, with no task- or outcome-based incentives.

Despite these findings, the development sector typically operates to the contrary, with many overseers in the industry focusing precisely on measuring the measurable components

of quality. The unsurprising effect of this practice is to over-incentivize production along the measureable dimensions of quality relative to the unobservable dimensions. The prime example of this problem is so-called “results-based management,” a framework that is used to run and evaluate many aid programs. Since results must be observable to make management results-based, these frameworks are predicated on rewarding the measurable aspects of aid delivery. Overall, the message of results-based management is to achieve the target at all costs, often at the expense of the higher-order components of quality.

From my experience in the development sector, a most illustrative example of this dynamic occurred when the Uganda Ministry of Health established the percentage of women completing all four recommended antenatal care visits as a key performance indicator for district health offices. In response, a local health officer, under pressure from the target, decided to stipulate that mosquito bednets only be provided to pregnant women at the fourth antenatal care visit, as a way to motivate the women to complete all of the visits. Demonstrating the pitfalls of results-based management, this approach wholly undermined the purpose of the bednets, which were meant to protect a woman from malaria during as much of the gestational period as possible.<sup>1</sup>

While the foregoing discussion primarily addresses quality in the context of contractual relationships, a separate, albeit related, matter is the quality of firms as standalone entities.<sup>2</sup> Specifically, how can we objectively evaluate the quality of a firm on its own? This question is particularly challenging in the development sector, given the complex nature of aid. To understand at least some of what makes the development sector more complicated, it is useful to draw a comparison to the business sector. Those in the business of making money have a straightforward and objective metric of success: profit. Thus, from the standpoint of those with interests in a firm, the business’s profit might be thought of as representing quality. In this context, profit provides a clear, precise measure of quality. As a firm-level measure, profitability also creates incentives that flow down through the firm and shape

---

<sup>1</sup>What was more, the bednets were provided by a third-party organization, so the health officer should not have even been able to exercise so much authority over them. This observation demonstrated the fact that the impact of external resource provision is highly dependent on the systems that receive the resources.

<sup>2</sup>Of course, contracts are ubiquitous, and it may be technically impossible to evaluate firms in a setting independent of any contracts. For instance, even the relationship between a firm and its owners may be viewed as a principal-agent contract. Nevertheless, the firm-only viewpoint is informative.



the operations of its components. Moreover, measures of profit transcend businesses of all industries, providing a simple, universal metric to enable valuation across firms even when the underlying substance differs. Profitability also plays an important dynamic role: unprofitable — and therefore low-quality — firms are ultimately driven out of business.<sup>3</sup>

In contrast to the for-profit sector, the development industry, and likely the non-profit sector as a whole, has no simple metric of quality. Following from the example of quality measurement for for-profit firms, the ideal analogy in the development sector would be to measure the quality of organizations based on their impact on social welfare, since their stated goal is to improve society. Indeed, this analogy lies at the heart of the results-based management approach. Such frameworks are certainly well-intended — for measuring ultimate results would be an ideal way to both evaluate and incentivize quality.<sup>4</sup> In practice, however, it is virtually infeasible in the aid industry to implement results-based frameworks as intended. Beside the aforementioned incentive problems associated with prioritizing what is measurable, the very task of defining results also faces inherent difficulties. In fact, the term “results” in results-based management tends to be a misnomer. Instead of measuring true results — such as the impact of a health program on mortality — results-based frameworks often measure inputs. For instance, typical “outcomes” under results-based management might include the number of children covered by an intervention or the number of mosquito bednets distributed in a given area. While such measures may certainly play a role in assessing performance, they tell us little about ultimate impact, a key question since the translation of input into impact is rarely straightforward. The measures are also incomplete,

---

<sup>3</sup>To be clear, the use of the term “quality” in this context is with respect to the overall objectives of the firm and those with interests in it. Quality here may not be synonymous with the firm’s contribution to social welfare. However, in an ideal setting, a sound regulatory environment would indeed align the firm’s profit-making incentives with social welfare. Regardless, the main point of the discussion is that profit provides a strong measure of firm-level quality within the context of the for-profit market.

<sup>4</sup>Even though I have drawn a distinction between a contract-related quality perspective and a firm-only quality perspective, it is important to note that one reason why results-based frameworks are so common in the development sector likely relates to the prominence of contractual relationships throughout the industry. As part of these contracts, aid donors institute results-based management systems in an attempt to ensure good performance by aid organizations. This operating arrangement is related to another dynamic looming underneath many of the quality issues addressed herein: that aid organizations tend to be highly dependent on external funding. However, even if more aid organizations were self-sustaining, profitability would not serve as a measure of their contribution to social welfare. Most of all, whether quality is viewed from the standpoint of an organization itself or that of its donors, the fundamental question of how quality should be defined remains essential.

at least partly due to the unobservable dimensions of quality. With these observations in mind, other important quality-related questions that we might want to know include the following: For the children, what was the nature of the intervention coverage? Were the delivered services of high quality? For the mosquito bednets, how do we know that the items reached the intended end users? Were the bednets instead siphoned off at lower levels of the distribution chain? Were adequate monitoring mechanisms in place to prevent such diversion?

Most fundamentally, is it possible to consistently and reliably measure impact in terms of social welfare? For instance, for a health organization, how is its impact on health to be measured? If we can even determine how best to measure the potential impact, how confidently can we then attribute observed changes in health to the organization? Alternatively, what about an organization that works on health, education, and environmental management? How is its impact across these disparate areas to be aggregated and assessed?

All of the foregoing discussion can be distilled to two main points. First, the act of measurement itself is bound to have the unintended consequence of shifting production incentives away from unobservable quality toward that which can be measured. Second, the choice of measurement, particularly in terms of capturing impact, is non-trivial. While the questions posed by these observations are difficult, they are not entirely answerless. For instance, to deal with these problems, an entire monitoring and evaluation industry, dedicated to the measurement of quality, has arisen. In addition, at least for particular programs, if not for entire organizations, we can apply rigorous (e.g., randomized) evaluation methods to causally attribute observed effects. Yet, many of these efforts fall short in accomplishing their stated goals — most of all because they do not address the crux of the quality problem in the development sector. Often, for example, monitoring and evaluation activities measure the wrong thing, such as the number of mosquito bednets distributed. In addition, monitoring and evaluation tends to be biased, and data are frequently manipulated, given the pressures on organizations to “produce results.” At the same time, rigorous evaluations are challenging and costly to scale up, due to the level of expert academic input required and the intensive nature of such studies. Moreover, and most critically, proper incentives must exist to motivate the performance of rigorous evaluations in the first place.

Much like the health service quality problems addressed in Chapter 2, one possible solution might be to invest more heavily in training for the conduct of sound evaluation. However, such approaches are bound to make little difference without fundamental change to the institutional structures and incentives in the development sector. Indeed, my observations show that substandard service delivery by the aid industry is not for lack of skill. Absent change to the underlying incentives, we will be left with a fragmented industry that has no simple way to assess its contribution to social welfare.

In large part, the development sector has been asking the wrong questions. Unfortunately, no matter how well we answer the wrong question, we will still arrive at a wrong answer. How, then, can we begin to ask the right questions? To address the foundational problems at play, we must entirely rethink our approach to quality in the development sector. Most importantly, we must accept that we will never identify an appropriate metric to use in measuring the quality of aid organizations, and we must therefore turn to next-best solutions. This recognition stems from the observation that development assistance is an inherently suboptimal solution — in fact, a third-best solution in response to government failure layered on top of market failure.

More specifically, without an objective outcome-based measure of quality, we should instead conceive of quality as being the product of properly aligned incentives. In this fashion, rather than focus on outcomes, we should focus on the upstream institutional structures that create quality. Even if we cannot use outcome measurement to directly assess the production of quality itself, we can use theoretical and empirical observation to evaluate the institutional incentives needed to produce quality. In essence, this perspective involves shifting our focus away from the “what” to the “how” — away from the product itself to the actual process of creating the product. This shift can also be termed as moving from “measurement accountability” to “structural accountability.” Although we will always have to employ some form of measurement, the purpose of structural accountability is to focus more on measuring underlying institutional quality rather than product quality.

To put these observations in concrete terms, such a perspective brings us to the core internal management operations of organizations. For example, how are staff members for an organization recruited to ensure that the best candidates receive jobs? How does the

organization incentivize existing employees to deliver the most for the communities they serve? According to what systems do employees work together to maximize the quality of output? How does the organization carry out its service delivery, from the very initial stages of needs assessment through on-the-ground implementation? What kinds of policies does the organization have in place to prevent corruption and wastage? How does the organization keep people at the center of all of its work and ensure that its initiatives remain directly aligned with the interests of the populations that it serves? In the following sections, while I will not answer all of these questions, I will offer some deeper ways for thinking about how to foster institutional quality related to these matters.

### 5.3 Quality Over the Production Cycle

The production cycle, or product development, is ubiquitous across many different tasks in life. In the development sector, where social services are the products, the production cycle can be used to characterize the process of program development and implementation. While the cycle will always be unique to the specific task or product in question, the program production cycle would generally consist of the following phases of activities: needs assessment and research, program conceptualization, design of implementation protocols, community-level preparation, and service delivery. For a for-profit firm, the production cycle would follow an analogous series of phases — for example, market and product research, product design and development, manufacturing, sales and marketing, and product delivery. We might also think of field research projects in terms of a production cycle consisting of the following stages: establishment of the research question, research instrument design and development, training of the field team, data collection, and analysis. In general, any task that can be broken into a series of stages leading to a final product can be characterized by a production cycle.

What can we learn about quality from the production cycle? Let me begin with a simple model. Suppose that a production cycle consists of  $N$  phases. In each phase, the producer must successfully complete the task associated with that stage of product development. For  $1 \leq n \leq N$ , let  $s_n$  be an indicator variable that equals 1 if the producer successfully completes

production phase  $n$  and 0 if the producer fails to complete the phase. The production cycle is path-dependent, with the successful completion of each stage depending on the successful completion of the previous stage. Specifically, the probability of failure in phase  $n$  is given by the following function, which we can refer to as the phase-specific failure rate:

$$P(s_n = 0) = \begin{cases} \varepsilon_n & \text{if } s_{n-1} = 1 \\ 1 & \text{if } s_{n-1} = 0 \end{cases} \quad (5.1)$$

In words, if the producer has successfully completed the preceding phase of production, then the producer faces an error rate of  $\varepsilon_n$ . As such, the error rate gives the probability of failure in the current phase  $n$ , while its complement,  $(1 - \varepsilon_n)$ , gives the probability of success in the phase. If, however, the producer has failed in the preceding phase of production, then the producer automatically fails the current stage as well. This feature means that a single error in any phase dooms the entire production process, conferring a feature to the production cycle that I refer to as *error propagation*. Note that, throughout all stages of the production cycle, the producer is never aware of the value of  $s_n$  and so does not know whether an error has been made in any phase. The producer continues with production regardless and only learns if the production process has failed to meet goals when the final product is brought to market.

Let us refer to  $\varepsilon_n$ , which is a component of the phase-specific failure rate, as the phase-specific error rate. As the production phases progress, suppose that the phase-specific error rates strictly decrease, such that  $\varepsilon_n > \varepsilon_{n+1}$  for all  $n$ . We can think of the declining phase-specific error rates as representing a narrowing production cycle. In particular, we might think of the first phase of production as encompassing a wide search interval for the “correct answer,” with substantial room for error. As production continues, the producer narrows down the space of possible products and, getting closer and closer to the final product, also experiences a lower error rate in each successive phase.

Now, rather than considering production phases on their own, let us integrate the phases together and assess the production cycle in a more comprehensive form. The phase-specific failure rate of equation (5.1) adopts the standpoint of phase  $n$ , conditional on the completion

of the previous phase of production. However, we can also evaluate failure rates for phase  $n$  from the outset of the production cycle, before any phases of production have been completed. I refer to these unconditional failure rates as total failure rates. Standing at the start of the production cycle, the successful completion of phase  $n$  depends both on its phase-specific error rate and on the successful completion of all previous stages. In short, there are two possibilities for failure in phase  $n$ : either (1) an error is made in phase  $n$  itself, or (2) an error is made in any one of the previous stages. The former scenario, which brings the phase-specific error rate into consideration, occurs if all previous stages have been completed successfully and the error is committed in phase  $n$ . The probability of such an event is given by the product of the probabilities of successfully completing each phase through  $n - 1$ , multiplied by  $\varepsilon_n$ . Alongside that possibility, the latter scenario means that we also must factor into the total probability of failure in phase  $n$  the possibility that an error has been made in any previous stage. In each of those stages, the possibility of failure consists of the same two foregoing scenarios. Thus, putting everything together, the total failure rate in phase  $n$  is equal to the sum of the probabilities of making an error in each phase leading up to and including phase  $n$ . We can write the total failure rate (TFR) as follows:

$$TFR_n = \begin{cases} \varepsilon_1 & \text{if } n = 1 \\ \varepsilon_1 + \sum_{j=2}^n \left[ \varepsilon_j \prod_{k=1}^{j-1} (1 - \varepsilon_k) \right] & \text{if } n > 1 \end{cases} \quad (5.2)$$

To compute the total failure rate for the entire production cycle, we can simply set  $n = N$ . As the total failure rate equation makes clear, the probability of failure in any given phase of production depends on the accumulation of error probabilities from all preceding phases, illustrating the property of error propagation through the production cycle. Notably, errors in earlier stages of production not only come with higher probabilities but also propagate through more phases.

Based on these observations, what does the model suggest? Most of all, the risk of error in earlier stages of production is more costly compared to the risk of error in later stages of production. In particular, correcting early-stage errors, which occur with greater likelihood, would necessitate redoing the production cycle from the beginning. Therefore,

the property of error propagation highlights the important role of “proactiveness,” implying that producers should invest heavily in “getting it right” early on. While in the present version of the model there is no role for firm investment, we can envision additional model features in which the error rates are not exogenous quantities but instead are a function of the producer’s choice of investment. With greater investment, the producer can reduce the error rates.

Along similar lines, the features of the production cycle should also motivate producers to closely monitor errors. Such monitoring would be aimed at overcoming the constraint in the present model that leaves producers unaware of whether errors have occurred until the very end of the production cycle. How can producers monitor errors? Namely, they can build in checkpoints at the end of each phase of production, where they rigorously review the task completed during that phase.

Roughly speaking, the idea of prospectively investing to reduce error rates corresponds to the notion of quality assurance, while the idea of creating checkpoints to monitor errors corresponds to the notion of quality control. Putting all of these observations together, we can then imagine that the input costs of production (including the costs of quality assurance and control), the cost of errors, and the expected product revenue will yield an optimal set of investment levels, corresponding to an optimized set of error rates, across all phases of production.

Lastly, note that this model of the production cycle exhibits significant flexibility, with applications at many levels and across multiple contexts. For example, based on notions of personnel quality, we could apply the model to a firm’s workforce, which I describe further in the following section. We can also apply the model at the level of specific projects, as highlighted in the introduction to this section. In addition, the model can be applied to particular components of projects — for example, if a product is comprised of multiple parts. Such an application means that we need not conceive of errors as dooming an entire product development process but can instead think of them as affecting specific product parts. Thematically, the model’s core emphasis on proactiveness captures Section 5.2’s discussion about asking the right questions. If the wrong question is asked from the starting phase of production, the error will propagate throughout the entire cycle. All in all, the

model offers insights for virtually any phase-based task.

## 5.4 Quality through Selection

Given the inherent difficulties of establishing an outcome-based measure of quality for the development sector, how can development organizations foster a high level of quality amongst their workforces? A key lesson of Section 5.2 was the critical need to adopt a more structural perspective on quality — one that looks more closely at the underlying systems and incentives in place to create quality than at final production outcomes. This lesson, combined with observations from Section 5.3, offers important insight for the essential operational area of personnel.

Let us first revisit Section 5.2’s comparison between development organizations and for-profit firms. In the for-profit sector, the ownership of firms has the favorable effect of (generally) aligning the interests of the firm’s owners with those of the firm. As stakeholders in a firm, the owners bear the firm’s risk and pay a personal price when the firm fails to meet its profitability goals — an arrangement that naturally incentivizes the owners to do what is best for the firm. In contrast, insofar as we conceive of the quality of development organizations as their contribution to social welfare, the leaders of such firms often bear minimal risk of the entities that they manage, since they typically suffer few consequences when their firms do not meet social welfare goals. In a more practical sense, since development organizations are not typically incentivized based on true measures of social welfare, the leaders do indeed bear some of the risk of the organizations; for instance, a firm’s failure to secure funding might result in job losses at the leadership level. Again, however, this risk-bearing generally is not aligned with real quality, bringing us back to the question of how quality ought to be measured.

In this setting, how can we ensure strong performance by the personnel of development organizations? If an organization’s contribution to social welfare could be easily measured, one option might be to incentivize the leaders based on such assessment. In practice, however, this approach is bound to run into many of the problems elucidated in Section 5.2. Even if such an approach were feasible at the level of senior leadership, it would likely be more



challenging for lower-level employees, since each employee's social value-added would have to be determined. With these difficulties in mind, I want to highlight the key, next-best approach: selection.

In the spirit of Section 5.3, we can conceive of personnel quality as the outcome of a production cycle. The first stage of this cycle would be a firm's recruitment process, including how widely the firm advertises for positions and which populations the firm targets for a given position. The next stage would be job interviews, including the rigor of the evaluation process. Once an employee is hired, the personnel quality production cycle stages include training and orientation activities, personnel rules and regulations, and compensation schemes, all of which are designed to maximize the quality of existing employees. Roughly speaking, we can think of this production cycle as a series of two main components: first, screening, and, second, regulation. The screening process takes place prior to hiring, in an attempt to select the best candidates, while the regulation process occurs after hiring, in an attempt to shape the chosen employees into the best performers possible. The production cycle is akin to a long funnel, beginning with a wide search interval that is repeatedly filtered through both screening and regulation with the hope of ultimately producing the organization's ideal employee.

While all stages of the personnel quality production cycle are important, the screening stages appear to be the most critical for building a quality workforce in the development sector. Without adequate measures available to properly incentivize performance, the next-best solution is to focus on the more upstream determinants of quality. In the workforce context, such upstream quality measures mean selecting the "right" people for the job. After all, as Section 5.3 shows, errors in the earlier stages of production are most costly. If an organization selects the wrong person for employment, the firm will never be able to transform this individual into the ideal employee. Although the organization's regulatory systems may be able to keep the employee's performance within certain parameters, the firm is bound to pay significant costs for landing on the wrong person and to regret that its screening procedures did not remove this individual from consideration early in the recruitment process. Altogether, while screening is essential for all firms, it is doubly important in the development sector that organizations wishing to produce quality invest heavily in the recruitment stages

of the personnel quality production cycle.

What do these observations mean in practice? While the answer differs depending on the context of each particular job, the general priority for development organizations must be to select employees who demonstrate traits such as the following: a positive attitude, passion for serving vulnerable communities, a hard work ethic, and a strong moral compass — all characteristics that tend to be related to the production of the unobservable and intangible components of quality.<sup>5</sup> In addition, organizations should instate rigorous, comprehensive screening procedures, such as aptitude exams, technical skill exams, essay interviews, interview panels, and multiple interview rounds. Although none of these recommendations may be especially groundbreaking, many development organizations follow recruitment procedures directly to the contrary. In fact, nepotism and bribery often play a key role in recruitment in the development sector, precluding the role of meritocratic processes in staff hiring.<sup>6</sup> Perhaps if we instituted more concerted efforts to address such corruption — if those evaluating the development sector focused more on the importance of rigorous internal management operations and less on the number of mosquito bednets distributed — we would be able to create a higher-quality aid industry.

## 5.5 Institutions as a Determinant of Labor Quality

While selection appears to be the most important ingredient of workforce quality for development organizations, proper selection by itself will hardly guarantee a high-performing team of personnel. As Section 5.4 explains, the personnel quality production cycle consists of two broad components: screening, followed by regulation. After the best possible candidates have been selected for jobs, the regulatory steps of the production cycle aim to optimize the behavior of the chosen workers. While much can be said about these regulatory stages, this section focuses on one regulatory aspect in particular: the systems according to which

---

<sup>5</sup>This conclusion relates to work on intrinsic motivation by [Prendergast \(2007\)](#), who shows that it is socially desirable for bureaucracies to select employees who have biased preferences relative to those of society. Accordingly, the preferences of some bureaucrats, such as social service providers, may be biased in favor of their clients. In contrast, other bureaucrats, such as tax collectors, may be intrinsically biased against the welfare of their clients.

<sup>6</sup>Many organizations will also claim to follow meritocratic procedures, but the documentation for such claims will be forged.

personnel carry out their work. Note, however, that the accompanying lessons go far beyond the personnel quality production cycle; they apply to any production cycle, since the systems governing labor input are primarily intended to minimize production error rates.

Let me begin with an example from my work with PHP in Uganda. In the course of carrying out both research and service activities, a key task in project development is almost always to translate implementation materials from one language to another. Over several years, PHP recognized that its translation process fell short of its standards — and, indeed, that the translation procedures employed in the development sector as a whole regularly failed to maintain the fidelity of the original text undergoing translation. For instance, common problems include inconsistent translation of key words or phrases, even though materials are supposed to be standardized; misinterpretation of the source text by translators; and poor rigor in segregating the roles of all of those involved in the translation process.<sup>7</sup> These problems often lead to grave but unrecognized mistakes — because if materials developed in one language cannot be properly translated into another language, the entire purpose of those materials is undermined. Thus, though often neglected, the task of translation is one of the most essential steps in quality program development.

Given the challenges of working across languages and the professional, standardized nature of many project materials, translation can be a complicated process. However, with some advanced planning, it does not have to be so. Consequently, within PHP, we undertook the task of developing what the organization terms its *Translation Protocol*, a 13-step procedure complete with guidelines and an accompanying checklist to govern the process that all staff members follow in performing translation work with one another. Overall, the *Translation Protocol* functions as a quality assurance tool, providing a highly systematic approach to translation procedures.

We implemented the protocol for the first time for the project activities of Chapters 1 and 2 of this dissertation. Its use was a resounding success. It was immediately evident that the final, translated products created by our team members were the highest quality materials they had ever developed. An external reviewer confirmed this fact when he was

---

<sup>7</sup>In general, high-fidelity translation must follow three primary steps: forward translation by one team, backward translation by another team that does not have access to the source text, and reconciliation of discrepancies by a third team.

equally impressed by the materials.

In addition to the *Translation Protocol*, we also developed a breadth of other program-specific protocols and checklists, as described in Chapter 2. The effect of these tools was to minimize production error rates over the course of all of our activities. Moreover, as resources like checklists cover many mundane and easily forgettable tasks, the collection of materials positioned staff members to focus more of their energy on the higher-order components of quality, while simultaneously ensuring that the staff consistently fulfilled the more basic quality components.

So what are the key takeaways from this case study? Most importantly, institutions — or the way in which we organize people to work together in teams — are a significant input into labor quality. More specifically, we can think of labor quality as the product of three inputs: (1) individual human capital, (2) worker complementarities, and (3) institutions. For the first input, each individual worker possesses a certain level of human capital. For the second input, when people work together, they create natural production complementarities, due to dynamics such as camaraderie and idea exchange. Finally, the third input of institutions, which provide a set of rules according to which people perform collaborative work, can significantly multiply the production possibilities from the first two inputs.

Indeed, what was remarkable about our work in Uganda was that the staff members performing the tasks largely came from the same team that had been working together for many years. Their individual human capital levels remained constant, as did the worker complementarities created between them. However, with additional structure and regulation of the team, we were able to dramatically improve performance.

In such team settings, we might think of institutional organization as the role of leadership. If leadership plays such an important role in producing favorable outcomes, how, then, can we foster better leadership on a wide scale? In the broader settings of social systems, that is the role of governance — for governance is the institutionalization of leadership itself. When effectively structured, governance systems produce the leaders that we need to solve social problems.

With regard to solving problems in global development, a common refrain heard about places such as Uganda is, “That can’t be done there.” Even many Ugandans themselves are

prone to make similar statements. However, our programmatic efforts serve as living proof that high-quality work can be performed in a low-resource environment. The overall lesson confirms the premise of Chapter 2: that with the right incentives in place, the Ugandan government can deliver high-quality health services. In the end, quality — whether in the for-profit sector, the public sector, or the development sector — always boils down to governance.

## References

- Holmstrom, Bengt, and Paul Milgrom. 1991. "Multitask Principal-Agent Analyses: Incentive Contracts, Asset Ownership, and Job Design." *Journal of Law, Economics, & Organization* 7:24.
- Prendergast, Canice. 2007. "The Motivation and Bias of Bureaucrats." *American Economic Review* 97 (1): 180–196.

# APPENDIX A

## Appendix for Chapter 1

### A.1 Methodological Specifications

#### A.1.1 Covariate Missing Values

For the balance tests shown in Table 1.1, missing values of covariates are coded as blanks. Therefore, observations with missing values for a given covariate are excluded from the respective different-in-means test. To account for this fact, I also test balance for the missing any covariate indicator, which equals one if an observation has a missing value for any of the other covariates and zero otherwise.

For the demographic controls included in the regression shown in the second column of Table 1.3, all missing values are coded as -9999 to enable observations with missing data to still be included in the regression procedure. For each specific covariate, I construct an additional missing indicator variable, which equals one if the corresponding covariate has the missing value of -9999 and zero otherwise. I then include all of the covariate missing indicators amongst the regression controls. These indicators serve to control for missingness, so that the missing data covariate values do not have a direct effect on the covariate coefficient estimates.

#### A.1.2 Computation of Economic Index

The covariates used for both the balance tests shown in Table 1.1 and the regression in the second column of Table 1.3 include an economic index variable that proxies for household economic status. I construct the economic index using principal component analysis (PCA) of twenty-one variables. PCA finds the orthogonal linear combinations of a set of

variables that best capture the overlapping information contained by the variables; the first principal component is the linear combination with maximum variance and is used as the economic index. See [Filmer and Pritchett \(2001\)](#) for more discussion of the method and for a demonstration of its validity in constructing such an index.

The variables in the index include the following: ratings of (1) the quality of house external wall construction material and (2) the quality of house floor construction material; indicators for ownership of each of the following by any household member: (3) bicycle, (4) motorcycle, and (5) car or truck; indicators for possession/ownership of each of the following: (6) non-solar electricity, (7) solar panel, (8) radio, (9) television, (10) mobile telephone, (11) refrigerator, (12) clock or watch, (13) bed/mattress, (14) chair/stool, (15) table, and (16) mosquito bednet; (17) number of acres of agricultural land owned by household members; and number of each of the following livestock animals owned by the household: (18) chickens, (19) goats, (20) pigs, and (21) cows.

For the variable measuring the quality of house external wall construction material, the rating is given a value of 1 for mud and pole; 2 for unburnt bricks with mud or unburnt bricks with cement; 3 for burnt/stabilized bricks; and 4 for cement blocks, concrete/stone, or wood. For the variable measuring the quality of house floor construction material, the rating is given a value of 1 for rammed earth; 2 for bricks or stone; 3 for cement screed, concrete, or wood; and 4 for tiles. The household survey also collected information on the construction material used for the roof, but this measure exhibited virtually zero variability across households and therefore was not included in the economic index.



## A.2 Intervention Implementation Materials

### A.2.1 Encouragement Message Verbal Scripts (English)

This section contains the respective message scripts that the baseline survey interviewers read to deliver the oral encouragement message to each household.

#### Neutral

I would like to inform you about an upcoming meeting between citizens and the LC3 Chairperson regarding government health services. The purpose of the meeting is to provide an opportunity for community members to discuss with the LC3 Chairperson any concerns or complaints related to the quality of health services at the local Health Center III. We encourage you to attend the meeting.

#### Public Duty

I would like to inform you about an upcoming meeting between citizens and the LC3 Chairperson regarding government health services. The purpose of the meeting is to provide an opportunity for community members to discuss with the LC3 Chairperson any concerns or complaints related to the quality of health services at the local Health Center III.

As a member of the community and as a citizen of Uganda, you have a duty to participate in the meeting for the sake of overall community welfare. Your participation in the meeting is vital, since citizens have a key role to play in the functioning of the local government. In fact, you have an obligation to your fellow citizens to attend the meeting and contribute your input. The viewpoints you share during the meeting will support improvements at the Health Center III, which provides health services to the entire community. Such improvements in health service delivery can, in turn, improve the health of everyone in the community. Therefore, we encourage you to attend the meeting to fulfill your responsibility as a member of this community to advocate for better quality health services.

## Private Benefits

I would like to inform you about an upcoming meeting between citizens and the LC3 Chairperson regarding government health services. The purpose of the meeting is to provide an opportunity for community members to discuss with the LC3 Chairperson any concerns or complaints related to the quality of health services at the local Health Center III.

If the meeting helps to improve local health services, you and your family will experience substantial benefits. Thanks to the improved quality of the Health Center III, you and your family will receive better services when you visit the health facility to seek care. In turn, this means that your health will improve, which will increase your work capacity at your home. In addition, if you receive better services from the Health Center III, this means you will spend less money and time on trying to obtain good health services. All of these benefits will help to increase your work productivity, leading to greater income and better development for your family. Therefore, we encourage you to attend the meeting to support your personal and family development.

## Civic Participation

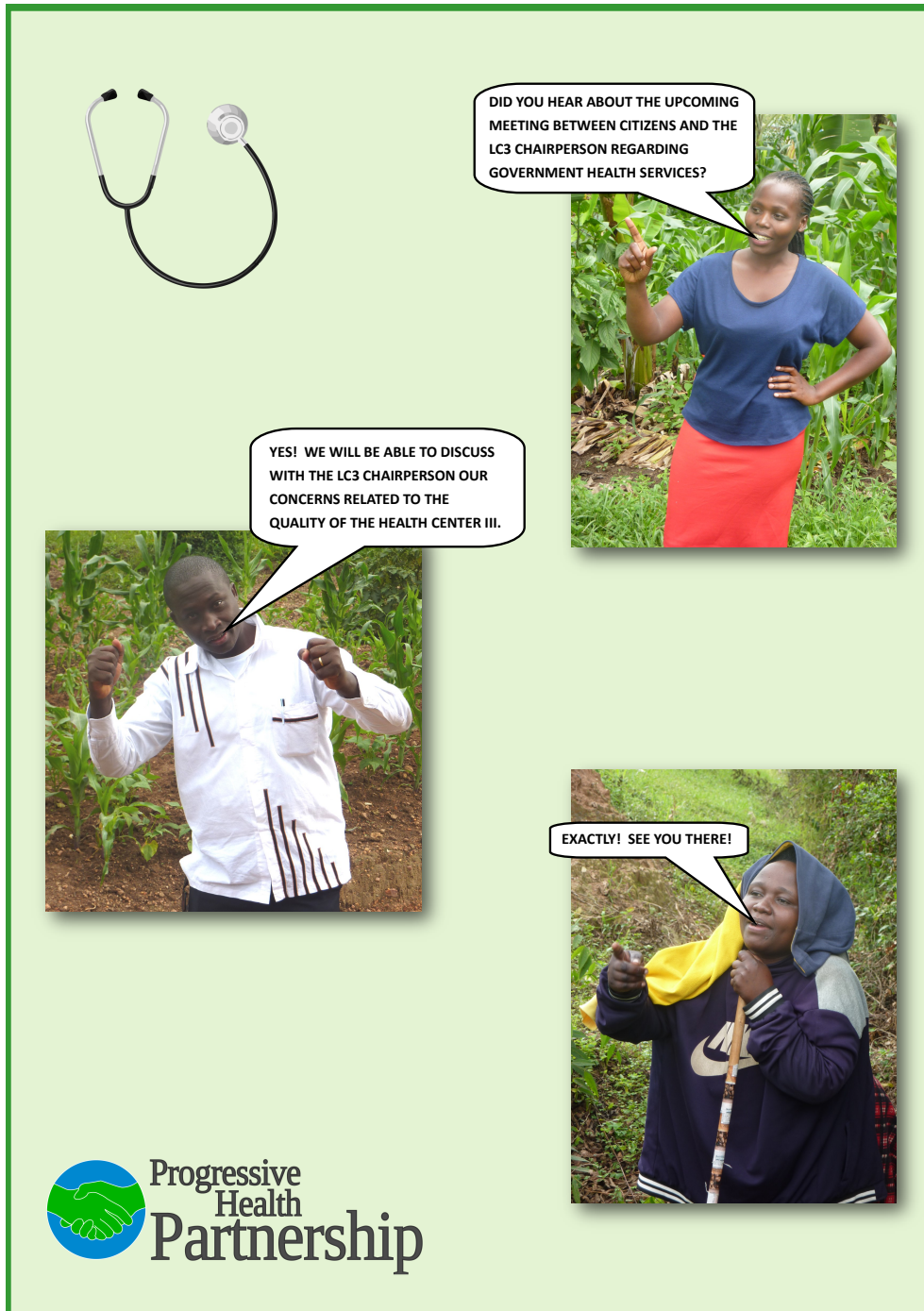
I would like to inform you about an upcoming meeting between citizens and the LC3 Chairperson regarding government health services. The purpose of the meeting is to provide an opportunity for community members to discuss with the LC3 Chairperson any concerns or complaints related to the quality of health services at the local Health Center III.

The meeting will specifically aim to give different social groups—such as poor families, women, the elderly, the youth, and disabled individuals—an opportunity to raise their voices with the LC3 Chairperson about the quality of health services at the local Health Center III. The viewpoints of many people are not always heard or listened to in the course of conducting community affairs. However, this meeting will specifically give people of different backgrounds and social groups a platform to speak and be heard. People who attend the meeting will be able to ask questions, raise concerns, and express complaints—and then receive answers. Everyone’s viewpoint will be respected. Therefore, we encourage you to attend the meeting to take advantage of this unique opportunity to raise your voice and to hear the perspectives of others.

## A.2.2 Encouragement Message Posters (English)

In addition to the verbal encouragement messages read to households, each household received a poster emphasizing the encouragement message theme. Copies of the posters are shown below.

### Neutral Poster



## Public Duty Poster

# Support your community by helping to improve health services!



DID YOU HEAR ABOUT THE UPCOMING MEETING BETWEEN CITIZENS AND THE LC3 CHAIRPERSON REGARDING GOVERNMENT HEALTH SERVICES?



YES! AS MEMBERS OF THE COMMUNITY, WE HAVE A DUTY TO DISCUSS WITH THE LC3 CHAIRPERSON OUR CONCERNS RELATED TO THE QUALITY OF THE HEALTH CENTER III.



EXACTLY! OUR PARTICIPATION IN THE MEETING CAN IMPROVE THE HEALTH AND WELFARE OF OUR ENTIRE COMMUNITY. I WILL EVEN INFORM MY FELLOW CITIZENS ABOUT IT. SEE YOU THERE!



Private Benefits Poster

# Good health promotes your family's development!



DID YOU HEAR ABOUT THE UPCOMING MEETING BETWEEN CITIZENS AND THE LC3 CHAIRPERSON REGARDING GOVERNMENT HEALTH SERVICES?



YES! BY DISCUSSING WITH THE LC3 CHAIRPERSON OUR CONCERNS RELATED TO THE QUALITY OF THE HEALTH CENTER III, WE WILL BE ABLE TO IMPROVE HEALTHCARE FOR OURSELVES AND OUR FAMILIES.



EXACTLY! IF YOU HAVE ACCESS TO GOOD HEALTH SERVICES, THEN YOU ARE ABLE TO WORK HARDER, SPEND LESS MONEY ON HEALTHCARE, AND EARN MORE INCOME FOR YOUR FAMILY. SEE YOU THERE!



## Civic Participation Poster

# Raise your voice with your leader!



DID YOU HEAR ABOUT THE UPCOMING MEETING BETWEEN CITIZENS AND THE LC3 CHAIRPERSON REGARDING GOVERNMENT HEALTH SERVICES?



YES! WE WILL HAVE AN OPPORTUNITY TO RAISE OUR VOICES WITH THE LC3 CHAIRPERSON ABOUT OUR CONCERNS RELATED TO THE QUALITY OF THE HEALTH CENTER III.



EXACTLY! THE VIEWPOINTS OF EVERYONE WILL BE RESPECTED—INCLUDING POOR FAMILIES, WOMEN, THE ELDERLY, THE YOUTH, AND THE DISABLED. I WILL ATTEND BECAUSE I WILL BE LISTENED TO. SEE YOU THERE!



### **A.2.3 Encouragement Message Information Sheets (English)**

Households also received an information sheet corresponding to their message arm, which served to provide a written version of the full oral message delivered to each household. Copies of the information sheets are provided below. The number in the lower left of the sheets provides the message group number, which was used to help interviewers ensure that they correctly matched the information sheet and the poster when distributing the materials to households. The information sheets also include the household ID in the lower right. The ID again helped ensure proper matching of materials by the interviewers. In addition, it is the information sheet that households were requested to bring with them to the meetings and submit, so that attendance could be directly verified based on household ID (see [Section 1.2.2](#) for further detail).



Neutral Information Sheet

**ANNOUNCEMENT OF HEALTH MEETING  
BETWEEN CITIZENS AND LC3 CHAIRPERSON**

Be sure to attend the upcoming meeting between local citizens and the LC3 Chairperson regarding government health services!

The meeting will be held at/near [MEETING LOCATION]. In the coming weeks and months, there will be announcements within your subcounty to inform you of the exact date and time of the meeting. Make sure to keep your ears and eyes open for the announcements so that you can get the information about when the meeting will be held.

1

[HOUSEHOLD ID]

## Public Duty Information Sheet

### ANNOUNCEMENT OF HEALTH MEETING BETWEEN CITIZENS AND LC3 CHAIRPERSON

Be sure to attend the upcoming meeting between local citizens and the LC3 Chairperson regarding government health services!

The meeting will be held at/near [MEETING LOCATION]. In the coming weeks and months, there will be announcements within your subcounty to inform you of the exact date and time of the meeting. Make sure to keep your ears and eyes open for the announcements so that you can get the information about when the meeting will be held.

#### Why Should You Attend?

- As a member of the community and as a citizen of Uganda, you have a duty to participate in the meeting for the sake of overall community welfare.
- Your participation in the meeting is vital, since citizens have a key role to play in the functioning of the local government.
- In fact, you have an obligation to your fellow citizens to attend the meeting and contribute your input.
- The viewpoints you share during the meeting will support improvements at the Health Center III, which provides health services to the entire community.
- Such improvements in health service delivery can, in turn, improve the health of everyone in the community.
- Therefore, we encourage you to attend the meeting to fulfill your responsibility as a member of this community to advocate for better quality health services.

2

[HOUSEHOLD ID]

## ANNOUNCEMENT OF HEALTH MEETING BETWEEN CITIZENS AND LC3 CHAIRPERSON

Be sure to attend the upcoming meeting between local citizens and the LC3 Chairperson regarding government health services!

The meeting will be held at/near [MEETING LOCATION]. In the coming weeks and months, there will be announcements within your subcounty to inform you of the exact date and time of the meeting. Make sure to keep your ears and eyes open for the announcements so that you can get the information about when the meeting will be held.

### Why Should You Attend?

- If the meeting helps to improve local health services, you and your family will experience substantial benefits.
- Thanks to the improved quality of the Health Center III, you and your family will receive better services when you visit the health facility to seek care.
- In turn, this means that your health will improve, which will increase your work capacity at your home.
- In addition, if you receive better services from the Health Center III, this means you will spend less money and time on trying to obtain good health services.
- All of these benefits will help to increase your work productivity, leading to greater income and better development for your family.
- Therefore, we encourage you to attend the meeting to support your personal and family development.

3

[HOUSEHOLD ID]

## Civic Participation Information Sheet

### ANNOUNCEMENT OF HEALTH MEETING BETWEEN CITIZENS AND LC3 CHAIRPERSON

Be sure to attend the upcoming meeting between local citizens and the LC3 Chairperson regarding government health services!

The meeting will be held at/near [MEETING LOCATION]. In the coming weeks and months, there will be announcements within your subcounty to inform you of the exact date and time of the meeting. Make sure to keep your ears and eyes open for the announcements so that you can get the information about when the meeting will be held.

#### Why Should You Attend?

- The meeting will specifically aim to give different social groups—such as poor families, women, the elderly, the youth, and disabled individuals—an opportunity to raise their voices with the LC3 Chairperson about the quality of health services at the local Health Center III.
- The viewpoints of many people are not always heard or listened to in the course of conducting community affairs.
- However, this meeting will specifically give people of different backgrounds and social groups a platform to speak and be heard.
- People who attend the meeting will be able to ask questions, raise concerns, and express complaints—and then receive answers.
- Everyone’s viewpoint will be respected.
- Therefore, we encourage you to attend the meeting to take advantage of this unique opportunity to raise your voice and to hear the perspectives of others.

4

[HOUSEHOLD ID]

#### **A.2.4 Encouragement Message Text Message Scripts (English)**

This section contains the text of the themed text messages that were sent to all households with telephone numbers on record.

##### **Neutral**

We are kindly reminding you to come for the community meeting regarding health services between the LC3 Chairperson and citizens on [DATE] at [TIME] at [LOCATION].

##### **Public Duty**

Support your community by helping to improve health services! We are kindly reminding you to come for the community meeting regarding health services between the LC3 Chairperson and citizens on [DATE] at [TIME] at [LOCATION].

##### **Private Benefits**

Good health promotes your family's development! We are kindly reminding you to come for the community meeting regarding health services between the LC3 Chairperson and citizens on [DATE] at [TIME] at [LOCATION].

##### **Civic Participation**


Raise your voice with your leader! We are kindly reminding you to come for the community meeting regarding health services between the LC3 Chairperson and citizens on [DATE] at [TIME] at [LOCATION].

### **A.3 Community-Wide Mobilization Materials**

As described in Section [1.2.1](#), PHP carries out community-wide mobilization in advance of each meeting. The following sections contain the key materials used for this process. All households, regardless of their message and survey assignment status, were exposed to the community-wide mobilization.

### A.3.1 Invitation Letters for Religious and Opinion Leaders (English)

Seven days prior to each meeting, PHP's local mobilizer delivers general invitation letters to various religious and opinion leaders in the target community. The letters serve to request their assistance in mobilizing community members for the meeting. A copy of the letter template is provided below.



---

Date: \_\_\_\_\_  
Name: \_\_\_\_\_  
Parish/Ward: \_\_\_\_\_  
Subcounty/Town Council/Division:  
\_\_\_\_\_

**RE: Invitation to participate in the Community-Centered Development Project meetings**

Dear Sir/Madam:

First of all, thank you for the great work you have done for the community.

On behalf of Progressive Health Partnership (PHP), a non-governmental organization, I would like to invite you to participate in the Community-Centered Development Project. Under this project, PHP will be organizing quarterly dialogue meetings between the LC3 Chairperson and the citizens in your community. The purpose of the meetings is to discuss concerns related to health services at the Health Center III.

We believe your support toward this activity will help enable the meetings to have the greatest impact possible on local healthcare delivery. Specifically, we kindly request that you help to inform community members of the upcoming meeting, which will take place as follows:

Date: \_\_\_\_\_  
Time: \_\_\_\_\_  
Location: \_\_\_\_\_

As a leader in the community, you play a critical role in local affairs. We therefore deeply value your participation in the dialogues and look forward to working together with you. Should you have any questions, please do not hesitate to contact me at [PHONE NUMBER] or [EMAIL ADDRESS].

Yours faithfully,

[NAME]  
[POSITION]  
Progressive Health Partnership

Progressive Health Partnership    Plot 14 Muti Drive, Boma, Mbarara    P.O. Box 1287  
Tel: +256-756-111-961    info@proghealth.org    www.proghealth.org

### A.3.2 Announcement Poster (English)

Seven days prior to each meeting, PHP's local mobilizer also hangs announcement posters throughout the community. The announcement poster format is provided below.

# HEALTH MEETING WITH LC<sub>3</sub> CHAIRPERSON

**Date:**

**Time:**

**Location:**

---

Join your fellow citizens for a meeting with the LC3 Chairperson to discuss the quality of service delivery at the Health Center III. Come to talk about your experiences and to express your concerns and complaints!

 Progressive Health Partnership

*For inquiry, please contact 0756-111-961.*



### **A.3.3 Text Message Reminder for Local Leaders (English)**

Both seven days and two days prior to each meeting, PHP sends text message reminders to all local leaders, in addition to calling them. A copy of the reminder text is provided below.

Come for the community meeting regarding health services between the LC3 Chairperson and citizens on [DATE] at [TIME] at [LOCATION]. Please inform your friends, relatives, neighbors, and colleagues. Don't miss out! For any inquiries, call [CONTACT PERSON NAME] at [PHONE NUMBER]. Thank you.

### **A.3.4 In-Field Mobilization Announcement Script (English)**

Both three days prior to and in the morning of the day of each meeting, PHP carries out in-field mobilization by driving throughout the target community with a mobile public address system to make announcements about the upcoming meeting. The text of the announcement script is provided below.

Progressive Health Partnership is an organization that works together with the government in health care service delivery to the people. We call everyone to be part of the community health dialogue together with the LC3 chairperson.

Venue: [VENUE]

Date: [DATE]

Time: [TIME]

We shall be discussing the quality of health care service delivery at the Health Center III. Please come and give your thoughts and concerns on how the community can work together to improve the services at the Health Center III.

## APPENDIX B

### Appendix for Chapter 2

#### B.1 Study Population of Localities

This section details how the population of localities — from which the sample localities were drawn — was defined. While the locality population included subcounties/town councils/divisions throughout the Ankole Region of Uganda, three modifications were made to the full list of localities in this region to establish the population. First, due to budgetary constraints, localities that were prohibitively far from PHP’s main office in Mbarara City were excluded from the population. Second, as some localities in the Ankole Region either had no Health Center III at all (contrary to government standards) or had higher-level government health facilities, they were also excluded to maintain uniformity of the local health service infrastructure vis-à-vis the study interventions and data collection. Third, due to an increasing trend of administrative unit division by the government in Uganda, the boundaries of some localities had changed since the 2016 election. Because boundary changes would likely render the 2016 margin of victory variable (which was used for the stratification) inaccurate, any localities that had experienced such changes were excluded from the population. Some of these units also included newly formed localities, which accounted for the absence of Health Center III’s in some cases.

#### B.2 Sampling Area Village Selection Criteria

As described in Section [2.4.3.2](#), the sampling area consisted of clusters of three villages within each parish/ward included in the study. The first village in each cluster was defined as the village containing either the Health Center III or the parish/ward administrative

headquarters, depending on the parish/ward in question. The remaining two villages were the nearest neighbors to the first village. The nearest neighboring villages were defined as the villages “of greatest accessibility” to the first village. Specifically, enumerators received the following guidelines to determine accessibility.

The term “of greatest accessibility” is intended to specify those villages that are most accessible to the Central Village. In determining the most accessible villages, you will need to exercise appropriate judgement. You should adhere to the following criteria, listed in order of importance, to determine accessibility.

1. *Time*: The length of an average trip from one village to the other should be the primary determinant of accessibility. Often, time will be related to the physical distance separating two villages. However, in some unusual cases—for instance, where the terrain is difficult or where the walking paths or roads are not direct—two villages could be closer to each other in terms of straight-line physical distance but further from each other in terms of time.
2. *Distance*: If judgements related to time are unclear, physical distance can help distinguish which villages are most accessible to the Central Village. The shorter the physical distance, the more accessible the village is.
3. *Borders*: Depending on the geographic shape of the villages, it is possible for multiple villages to border the Central Village. Some bordering villages may have longer shared borders than others. In general, villages with longer bordering areas should be considered more accessible to the Central Village.
4. *Population*: If after consideration of the above three criteria it remains difficult to differentiate between the level of accessibility of two or more villages, population should be used to “break the tie.” You may gather a rough estimate of the populations of each village and favor the village with greater population.

Upon meeting the LC1 Chairperson, use your skills to identify the two nearest neighboring villages, working to ensure that the LC1 Chairperson does not give you biased responses. For instance, in considering the abovementioned criteria, you may ask questions such as the following:

- Where do I go to next after this village? And then where?
- Which other villages are the greatest users of the Health Center III?
- If there is a meeting at the Parish/Ward Headquarters, residents from which villages are most likely to attend?

## B.3 Computation of Economic Index

The covariates used for both the predictors regression shown in Table 2.2 include an economic index variable that proxies for household economic status. I construct the economic index using principal component analysis (PCA) of twenty-one variables. PCA finds the orthogonal linear combinations of a set of variables that best capture the overlapping information contained by the variables; the first principal component is the linear combination with maximum variance and is used as the economic index. See [Filmer and Pritchett \(2001\)](#) for more discussion of the method and for a demonstration of its validity in constructing such an index.

The variables in the index include the following: ratings of (1) the quality of house external wall construction material and (2) the quality of house floor construction material; indicators for ownership of each of the following by any household member: (3) bicycle, (4) motorcycle, and (5) car or truck; indicators for possession/ownership of each of the following: (6) non-solar electricity, (7) solar panel, (8) radio, (9) television, (10) mobile telephone, (11) refrigerator, (12) clock or watch, (13) bed/mattress, (14) chair/stool, (15) table, and (16) mosquito bednet; (17) number of acres of agricultural land owned by household members; and number of each of the following livestock animals owned by the household: (18) chickens, (19) goats, (20) pigs, and (21) cows.

For the variable measuring the quality of house external wall construction material, the rating is given a value of 1 for mud and pole; 2 for unburnt bricks with mud or unburnt bricks with cement; 3 for burnt/stabilized bricks; and 4 for cement blocks, concrete/stone, or wood. For the variable measuring the quality of house floor construction material, the rating is given a value of 1 for rammed earth; 2 for bricks or stone; 3 for cement screed, concrete, or wood; and 4 for tiles. The household survey also collected information on the construction material used for the roof, but this measure exhibited virtually zero variability across households and therefore was not included in the economic index.

## B.4 The Ugandan Political Environment

This section provides a more detailed assessment of the likely impact of the meetings intervention in the context of the Ugandan political environment. In non-democratic settings, a fundamental question is whether the meetings intervention truly has the potential to impact health service provision. Indeed, it is possible that the success of the intervention may hinge on the functioning of local elections. If local electoral processes do not function well, the incentives of political leaders to adopt new behaviors in the health sector may be undercut. For instance, a couple recent studies in Uganda have observed effects that depend on underlying features of the local political environment ([Grossman and Michelitch 2018](#); [Raffler 2020](#)). Such conditioning interactions introduce another intellectually interesting dimension and underlie the stratified experiment design of this study.

In regard to Uganda specifically, several considerations suggest that the country's political environment provides an appropriate setting for the study. First, the Ugandan government has been described as a hybrid, semi-authoritarian regime ([Tripp 2010](#)). While Uganda does not exhibit democratic principles at the national executive level, citizens exercise substantial freedom in other areas of governance, and the country's local political systems operate according to different norms. Although evidence on local electoral outcomes is limited, local elections have been reported to be relatively competitive, with high turnover of local government councilors each cycle ([Awortwi 2010](#); [Kiyaga-Nsubuga and Olum 2009](#)).

Still, there is no doubt that local governance in Uganda, as in many low- and middle-income countries, faces a range of systemic weaknesses, which may introduce additional constraints on political processes and service delivery. However, many of these weaknesses constitute the very motivation for the citizen-politician meetings intervention. The intervention is directly aimed at strengthening democratic practices and may therefore serve as a possible counterbalancing strategy to address shortfalls in democratic functioning. While the underlying causes of status-quo behaviors such as clientelism and electoral malpractice are complex, one contributing factor is likely limited citizen participation in political processes. Without sufficient information on politician performance, for example, citizens may be constrained in making informed voting decisions and therefore may maximize utility merely

based on the clientelism of politicians.

In fact, such a dynamic has been observed in Benin by [Fujiwara and Wantchekon \(2013\)](#), who show that information-based campaigns — through candidate-endorsed town hall meetings discussing specific policy platforms — successfully reduce clientelism in presidential politics. Thus, the provision of performance information — implemented herein through the citizen-politician meetings — may serve to shift the status-quo equilibrium toward one in which citizens make decisions more informed by the realities of service delivery. Providing further evidence of this possibility in the setting of Uganda, [Larreguy et al. \(2017\)](#) describe focus group discussions that reveal that citizens have little information to assess candidates, with elected officials arguing that they are not responsible for service delivery and voters being uninformed about the appropriate parties to hold accountable for service quality problems. Results from the anti-vote buying campaign that the authors evaluate also confirm the plausibility of Ugandan citizens changing their voting behaviors (to vote for their preferred candidates) in response to information.

Lastly, as several studies in Uganda have pointed out, policy-focused citizen participation and consultation during non-electoral time periods is a persistent gap ([Kakumba 2010](#); [Devas and Grant 2003](#)). Such non-electoral participation would complement the role of elections, which tend to be fairly blunt accountability mechanisms. Pointing to the promise of the meetings intervention, [Grossman and Michelitch \(2018\)](#) observe positive effects on district politician performance of a transparency-scorecard intervention implemented in the middle of the electoral term. Notably, politicians may be driven by other forms of incentives in addition to electoral accountability. For instance, politicians are likely to be motivated by an aversion to public or professional humiliation. Even local political leaders who are confident in their electoral prospects are likely to prioritize positive relations with their constituents. Should the performance of political leaders fall short of citizen expectations, it is possible that leaders at higher levels of government will learn of the circumstances, providing an additional source of professional pressure on local politicians.

## B.5 Health Service Delivery in Uganda

### B.5.1 Health Sector Performance

This sub-section offers a more in-depth assessment of the state of health service delivery in Uganda around the time of the study. Some of the most recent evidence on the quality of health service provision comes from the [World Bank's \(2016\) \*Service Delivery Indicators\*](#) survey, which was carried out in 2013 across 387 health facilities throughout the country. The results of this survey demonstrate the persistence of alarming service delivery deficiencies. For example, health worker absenteeism rates exceed 47 percent in the public health sector, a level that suggests a decline in performance compared to the absenteeism rate of 37 percent reported several years earlier by [Chaudhury et al. \(2006\)](#). Moreover, health centers have on average only 52.3 percent of all essential drugs, 39.3 percent of essential maternal drugs, and 42.3 percent of essential pediatric drugs in stock. Results for vaccine levels are more favorable, with an average of 76.3 percent of all essential vaccines in stock at public health centers. With regard to equipment availability, only 39.0 percent of public health centers meet minimum requirements ([World Bank 2016](#)).

Notably, [Dizon-Ross, Dupas, and Robinson \(2017\)](#) present evidence that counters an extended series of studies documenting low health service quality in countries such as Uganda. Focusing on programs that distribute free malaria bednets, the authors find relatively high performance in the delivery of bednets by public sector health workers. However, as the authors acknowledge, the unique features of bednet programs — as publicly visible campaigns distributing a simple, preventative product for which willingness-to-pay tends to be low — may explain the low levels of corruption observed. Such features may not generalize to the broad set of health facility practices and outcomes that this study examines. For example, while the lack of a high-turnover retail market largely precludes diversion of bednets, diversion of inputs such as drugs and supplies is one of the key outcomes of interest in the setting of the present study.

Interestingly, while [Dizon-Ross, Dupas, and Robinson \(2017\)](#) report health worker absenteeism rates in a similar range as those cited above, they present additional evidence



showing that health worker performance appears to be high for tasks other than bednet distribution as well. However, the data employed here are restricted to a limited number of antenatal care indicators. Other process quality measures for antenatal care, such as blood pressure measurement and deworming, remain low in the Ankole Region ([UBOS and ICF 2018](#)). More generally, while antenatal care quality has been the emphasis of many health-care improvement programs, other aspects of healthcare — such as gestational age at first antenatal care visit, postnatal care utilization, and vaccination rates — continue to show substantial shortfalls, as described in the following sub-section.

### **B.5.2 Healthcare Delivery Gaps, Potential Health Impact Pathways, and Health Outcomes of Interest**

This sub-section establishes a more detailed health outcomes framework based on the key healthcare delivery gaps that the interventions are most likely to address. Extending from the substandard health facility performance indicators detailed in the preceding sub-section, Uganda continues to fall short in several aspects of health intervention delivery to patients. For instance, even though antenatal care delivery has improved substantially in recent years, women in Uganda present for their first antenatal care visit at a median gestational age of 4.7 months — implying several months of missed opportunities for health prevention measures. In addition, well over half of women and newborns in the Ankole Region of Uganda go without a single postnatal care visit. Vaccination rates remain relatively low, with 61.8 percent of children age 12-23 months having all basic vaccinations and only 34.4 percent having all age-appropriate vaccinations ([UBOS and ICF 2018](#)). Importantly, these low levels of intervention adoption and delivery reflect not only supply-side service gaps but also perceived low returns to seeking healthcare, which likely arise at least in part from poor service delivery itself. Along similar lines, recent public health literature has placed a strong emphasis on the role of respectful care in influencing utilization of health services ([Kyaddondo et al. 2017](#); [Sharma et al. 2015](#); [Abuya et al. 2015](#))

Following these observations, the interventions in this study have the potential to affect outcomes on both the supply and demand sides. On the supply side, the programs may

address two main dimensions of health facility performance: the intensive margin and the extensive margin. The intensive margin includes areas such as health worker absenteeism, drug and supply diversion, facility condition and sanitation, and respectful care, all of which local leaders can help to address. The extensive margin of facility performance includes the total capacity of local health centers, such as available infrastructure and staffing. If, as a result of either program, local politicians can successfully report and advocate upwards for increased resource allocation to their communities, the interventions can plausibly lead to extensive margin changes such as staff increases and infrastructural improvements.<sup>1</sup> More specifically, through the training program, political leaders gain skills to address the full breadth of potential supply-side problems at health facilities. In the meetings program, citizen complaints to politicians tend to relate more to the intensive margin of facility performance but can also relate to the extensive margin.

On the demand side, service delivery improvements due to either program may lead to increased healthcare utilization, helping to drive increased health intervention adoption and delivery. Because the meetings intervention in particular fosters the civic engagement of citizens to improve healthcare in their communities, community members in this arm may be especially likely to exhibit higher levels of healthcare utilization.

Based on the foregoing possible pathways, the study's health services outcomes fall into one of four categories, as highlighted in Section 2.4.3.1: (1) healthcare utilization, (2) healthcare quality, (3) intervention delivery and adoption, and (4) patient satisfaction. The healthcare utilization category includes measures such as number of visits to the local health center in the past 12 months, visits to other health providers, antenatal care completion by pregnant women, and postnatal care attendance and completion. The healthcare quality category includes measures such as drug and supply availability, health worker absence rate, wait time, use of equipment, facility condition and sanitation, and size of staff. The intervention delivery and adoption category includes childhood immunizations and treatment of common childhood illnesses. Finally, the patient satisfaction category includes measures of patient experience at recent health visits (e.g., respect of health workers, understanding of

---

<sup>1</sup>Should these changes be realized, however, they are likely to occur over an extended period of time and therefore fall within the scope of a longer-term study.

instructions) and overall satisfaction with the health center services. At the impact level, beyond these intermediary health outcomes, I examine child mortality, weight-for-age, and mid-upper arm circumference.

# APPENDIX C

## Appendix for Chapter 3

### C.1 Methodological Specifications

#### C.1.1 Covariate Missing Values

For the predictor regressions shown in Table 3.13, all missing values are coded as -9999 to enable observations with missing data to still be included in the regression procedure. For each specific covariate, I construct an additional missing indicator variable, which equals one if the corresponding covariate has the missing value of -9999 and zero otherwise. I then include all of the covariate missing indicators amongst the regressors. These indicators serve to control for missingness, so that the missing data covariate values do not have a direct effect on the covariate coefficient estimates.

#### C.1.2 Computation of Asset Index

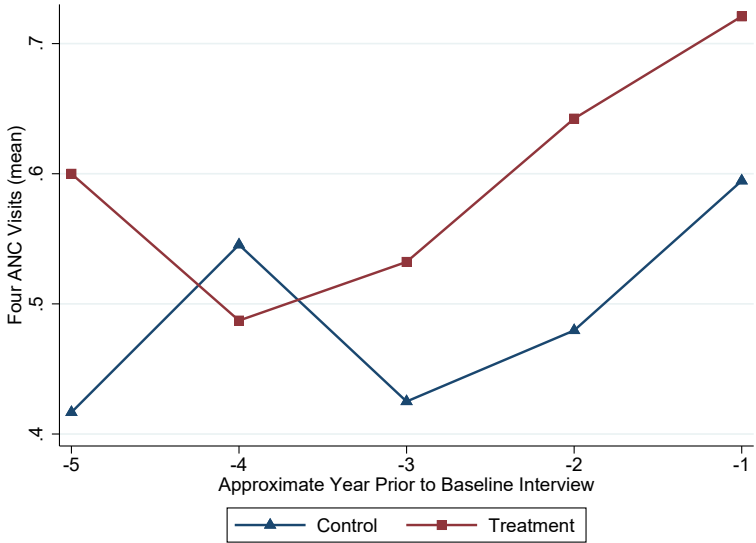
The covariates used for both the predictor regressions in Table 3.13 and the balance tests shown below in Table C.5 include an asset index variable that proxies for household economic status. I construct the asset index using principal component analysis (PCA) of fourteen variables. PCA finds the orthogonal linear combinations of a set of variables that best capture the overlapping information contained by the variables; the first principal component is the linear combination with maximum variance and is used as the economic index. See [Filmer and Pritchett \(2001\)](#) for more discussion of the method and for a demonstration of its validity in constructing such an index.

The variables in the index include the following: indicators for ownership of each of the following by any household member: (1) bicycle and (2) motorcycle or motor scooter;

indicators for possession/ownership of each of the following: (3) electricity, (4) radio, (5) mobile telephone, (6) clock or watch, (7) chair/stool, and (8) table; (9) number of hectares of agricultural land owned by household members; and number of each of the following livestock animals owned by the household: (10) chickens, (11) goats, (12) sheep, (13) pigs, and (14) cows.

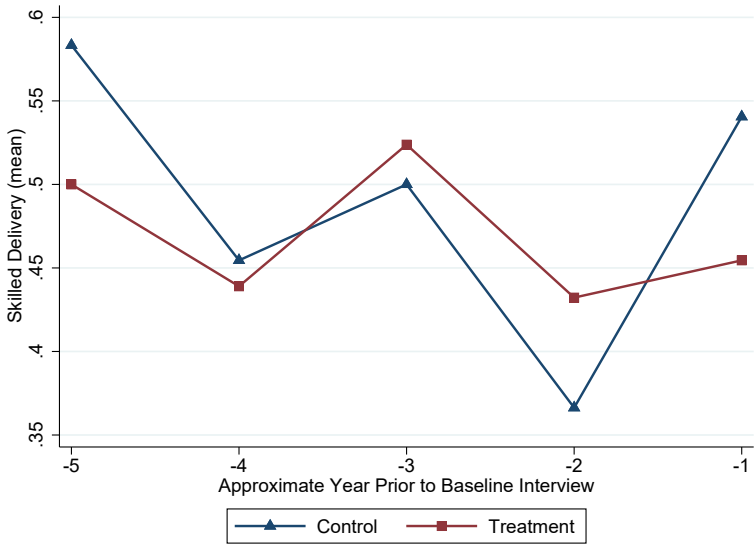
## C.2 Graphs of Pre-Program Outcome Trends

FIGURE C.1: PRE-TREATMENT FOUR ANC VISITS TRENDS



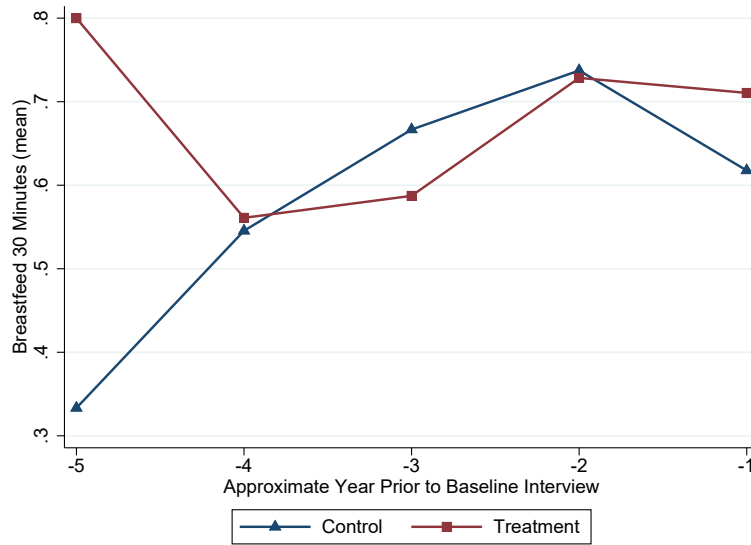
*Notes:* This figure shows pre-treatment trends in the control and treatment groups for the four antenatal care visits outcome variable. The trends are confined to the balanced panel of non-attrited respondents and, as described in Section 3.3.4, consist of repeated cross-sectional means from the respective subset of women in the sample who gave birth in each year prior to the baseline survey.

FIGURE C.2: PRE-TREATMENT SKILLED DELIVERY TRENDS



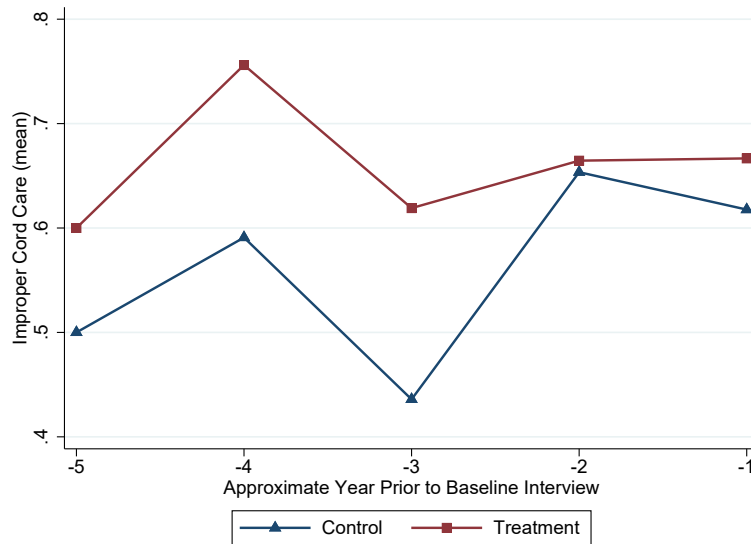
*Notes:* This figure shows pre-treatment trends in the control and treatment groups for the skilled delivery outcome variable. The trends are confined to the balanced panel of non-attrited respondents and, as described in Section 3.3.4, consist of repeated cross-sectional means from the respective subset of women in the sample who gave birth in each year prior to the baseline survey.

FIGURE C.3: PRE-TREATMENT BREASTFEED 30 MINUTES TRENDS



*Notes:* This figure shows pre-treatment trends in the control and treatment groups for the breastfeed within thirty minutes of birth outcome variable. The trends are confined to the balanced panel of non-attrited respondents and, as described in Section 3.3.4, consist of repeated cross-sectional means from the respective subset of women in the sample who gave birth in each year prior to the baseline survey.

FIGURE C.4: PRE-TREATMENT IMPROPER CORD CARE TRENDS



*Notes:* This figure shows pre-treatment trends in the control and treatment groups for the improper umbilical cord care outcome variable. The trends are confined to the balanced panel of non-attrited respondents and, as described in Section 3.3.4, consist of repeated cross-sectional means from the respective subset of women in the sample who gave birth in each year prior to the baseline survey.

## C.3 Additional Pre-Program Trend Tests

### C.3.1 Pooled Pre-Program Falsification Tests

In addition to the falsification tests between consecutive years of most recent births (shown in Section 3.3.4), I perform a pooled version of the falsification tests. For this specification, I pool the sub-samples for  $t = -4$  and  $t = -3$  and for  $t = -2$  and  $t = -1$  and then perform the placebo test across these two grouped time periods. This specification has the advantage of larger sample sizes, helping to verify that the results shown in the main text are not driven by small sample sizes. Table C.1 shows the results from the pooled test for the main specification, the CHW gender specification, and the CHW location specification. The interaction terms do not have a statistically significant effect for any of the outcome variables and specifications in this formulation of the test.

TABLE C.1: POOLED PRE-PROGRAM FALSIFICATION TESTS

	(1)	(2)	(3)	(4)
	Four ANC Visits	Skilled Delivery	Improper Cord Care	Breastfeed 30 Minutes
<i>Panel A: Main Specification</i>				
Post * Treatment	0.102 (0.090)	0.018 (0.090)	-0.161 (0.099)	0.064 (0.096)
<i>Panel B: CHW Gender Specification</i>				
Post * Treat-Male	0.098 (0.099)	0.052 (0.105)	-0.153 (0.103)	0.065 (0.104)
Post * Treat-Female	0.102 (0.125)	-0.038 (0.104)	-0.164 (0.121)	0.069 (0.117)
<i>Panel C: CHW Location Specification</i>				
Post * Treat-Home	0.160 (0.100)	0.034 (0.110)	-0.173 (0.108)	0.063 (0.112)
Post * Treat-Neigh	0.036 (0.107)	0.012 (0.098)	-0.141 (0.107)	0.072 (0.105)

*Notes:* This table displays the results of pooled pre-program falsification tests. Heteroskedasticity-robust standard errors, adjusted for clustering at the village level, are shown in parentheses. \*Significant at 10% level, \*\*Significant at 5% level, \*\*\*Significant at 1% level.



### C.3.2 Test of Equality in Linear Time Trends Between Treatment and Control

In addition to the placebo tests shown in both the main text and the appendix, I regress each outcome variable on time (i.e., years prior to baseline) to estimate the linear time trends in the treatment and control groups. Table C.2 displays the results for the main specification, showing the pre-program time trends for the full sample and treatment and control groups. The  $p$ -values in the table come from  $\chi^2$  tests for differences between the treatment and control group trends. None of the tests produce statistically significant results.

Tables C.3 and C.4 show linear time trends results for the CHW location and CHW gender specifications, respectively. The  $p$ -values in these tables come from  $\chi^2$  tests for differences between the control group trends and corresponding treatment group trends. Again, none of the tests produce statistically significant results.

TABLE C.2: PRE-PROGRAM TIME TRENDS ACROSS TREATMENT AND CONTROL GROUPS

	(1)	(2)	(3)	(4)
	Full Sample	Control	Treatment	$p$ -value
Four ANC Visits	0.048	0.029	0.065	0.358
Skilled Delivery	-0.015	-0.017	-0.014	0.927
Improper Cord Care	0.014	0.044	-0.008	0.220
Breastfeed 30 Minutes	0.053	0.065	0.043	0.591

*Notes.* This table displays the linear time trends of the outcome variables prior to the baseline survey, along with  $p$ -values for difference-in-means tests between the treatment and control groups. All tests use heteroskedasticity-robust standard errors adjusted for clustering at the village level.

TABLE C.3: PRE-PROGRAM TIME TRENDS BASED ON CHW LOCATION

	(1)	(2)	(3)	(4)	(5)	(6)
	Full Sample	Control	Treat-Home	Treat-Neigh	<i>p</i> -value (Home)	<i>p</i> -value (Neigh)
Four ANC Visits	0.048	0.029	0.090	0.040	0.234	0.813
Skilled Delivery	-0.015	-0.017	0.008	-0.035	0.642	0.690
Improper Cord Care	0.014	0.044	-0.035	0.018	0.154	0.580
Breastfeed 30 Minutes	0.053	0.065	0.048	0.039	0.736	0.568

*Notes.* Based on the treatment groups defined by CHW location, this table displays the linear time trends of the outcome variables prior to the baseline survey. The *p*-values correspond to difference-in-means tests between each treatment group and the control group. All tests use heteroskedasticity-robust standard errors adjusted for clustering at the village level.

TABLE C.4: PRE-PROGRAM TIME TRENDS BASED ON CHW GENDER

	(1)	(2)	(3)	(4)	(5)	(6)
	Full Sample	Control	Treat-Male	Treat-Female	<i>p</i> -value (Male)	<i>p</i> -value (Female)
Four ANC Visits	0.048	0.029	0.049	0.096	0.654	0.256
Skilled Delivery	-0.015	-0.017	0.016	-0.072	0.485	0.286
Improper Cord Care	0.014	0.044	-0.015	0.012	0.227	0.586
Breastfeed 30 Minutes	0.053	0.065	0.022	0.093	0.322	0.599

*Notes.* Based on the treatment groups defined by CHW gender, this table displays the linear time trends of the outcome variables prior to the baseline survey. The *p*-values correspond to difference-in-means tests between each treatment group and the control group. All tests use heteroskedasticity-robust standard errors adjusted for clustering at the village level.

## C.4 Analysis of Pre-Program Balance

Balance between the treatment and control groups is not a necessary condition for the difference-in-differences identification strategy. However, balance may be of interest insofar as the production of a treatment effect is a function of baseline outcome variable levels, rather than, say, additively separable.<sup>1</sup> Therefore, to supplement the pre-treatment trends analysis, I also evaluate balance between the treatment and control groups according to a variety of baseline dimensions. Table C.5 displays the results.

Out of twenty-two variables, the two groups demonstrate imbalance for only three. Nevertheless, a joint test of differences, shown at the bottom of the table, is statistically significant at the 5 percent level. Of the four outcome variables, only one — antenatal visit completion — is imbalanced across the groups, with the treatment group showing a higher baseline rate of antenatal visit completion. On the whole, potential concerns regarding balance provide greater justification for the difference-in-differences approach, since the identification strategy does not rely on such an assumption.

For the variables included in the balance tests, missing values are coded as blanks. Therefore, observations with missing values for a given covariate are excluded from the respective different-in-means test. To account for this fact, I also include a balance test for the missing any covariate indicator, which equals one if an observation has a missing value for any of the other covariates and zero otherwise.

---

<sup>1</sup>Additive separability of the treatment effect is not a formal assumption of difference-in-differences. However, the parallel trends assumption loosely implies this parametric paradigm, which may or may not be reasonable depending on the circumstances.

TABLE C.5: ANALYSIS OF PRE-INTERVENTION BALANCE ACROSS TREATMENT AND CONTROL GROUPS

	(1)	(2)	(3)	(4)
	Full Sample	Control	Treatment	<i>p</i> -value
Respondent Education Level	5.793	5.915	5.707	0.401
Household Head Education Level	6.240	6.074	6.357	0.381
Respondent is Household Head	0.026	0.023	0.028	0.688
Asset Index	0.025	0.041	0.013	0.857
Frequency of Listening to Radio	1.871	1.881	1.864	0.855
Household Owns Mosquito Bednet	0.786	0.769	0.798	0.427
Family Member Maternal/Neonatal Death	0.435	0.465	0.415	0.195
Number of Previous Pregnancies	3.165	3.132	3.188	0.749
Gestational Age (weeks)	18.452	18.597	18.351	0.677
Walked to Health Center	0.594	0.564	0.615	0.444
Accompanied by Husband to Health Center	0.123	0.155	0.100	0.050
Time to Health Facility	68.126	73.792	64.181	0.045
Burden to Come to ANC	0.503	0.514	0.496	0.638
Received ANC from Other Source	0.176	0.193	0.165	0.424
Has Taken Herbs	0.509	0.497	0.518	0.595
Number of HIV Prevention Methods Known	1.664	1.680	1.652	0.681
Husband Makes Health Decisions	0.427	0.439	0.419	0.609
Four ANC Visits	0.562	0.493	0.610	0.015
Skilled Delivery	0.451	0.443	0.457	0.789
Breastfeed 30 Minutes	0.670	0.660	0.677	0.687
Improper Cord Care	0.636	0.591	0.666	0.120
Missing Any Covariate	0.420	0.416	0.423	0.874
Joint test ( $\chi^2$ )				0.023

# APPENDIX D

## Appendix for Chapter 4

### D.1 Data Specifications

#### D.1.1 Data, Demographics, and Time Period

All of our computations using the ACS data occur at the individual level and therefore employ the person weights provided by the ACS. All analysis is restricted to the non-Hispanic population — that is, non-Hispanic Black and White men. Our estimation of the main model parameters uses 2014 data from the ACS and NVSR. Although U.S. life tables are currently available through 2018, ACS data are only available through 2015. In addition, in 2015, the ACS stopped reporting the “single race” variable, instead collecting data on race as a multiple-response variable. This change makes it difficult to merge ACS data from 2015 and afterward with NVSR mortality data, which are in fact reported based on single race. Altogether, these factors make 2014 the most recent, suitable year to use for the analysis.

To estimate  $\frac{\partial \Delta}{\partial s}$ , and  $\frac{\partial^2 \Delta}{\partial s^2}$  in the  $m(t, s)$  version of the model, we pool the ACS data from 2006 to 2014 to mitigate small-sample problems. We choose 2006 as the starting year because the ACS did not record institutionalization (group quarters) status from 2001 to 2005.

#### D.1.2 Imputation of Education Years

Using the *educd* variable in the ACS data, we impute years of education. While some of the *educd* codes correspond directly to specific grade levels, others do not and therefore require imputation. With a few small adjustments necessary for the ACS data, we follow the recommendations of [Jaeger \(1997\)](#) and [Jaeger \(2003\)](#), which address the imputation

of education years from *Current Population Survey* data. Specifically, Tables D.1 and D.2 detail our imputation methods. In 2008, the ACS changed the coding of the *educd* variable. Therefore, as the tables show, our imputation process differs for the periods before and after this change. Note that any grade levels not listed in the tables are recorded as exact values under the *educd* variable and therefore require no imputation.

TABLE D.1: IMPUTED EDUCATION YEARS FOR ACS DATA FROM 2008 AND AFTER

<i>educd</i> Code	Imputed Value
Nursery school, preschool	0 years
Kindergarten	0 years
GED	12 years
Some college, but less than 1 year	12 years
1 or more years of college credit, no degree	13 years
Associate's degree, type not specified	14 years
Master's degree	18 years
Professional degree beyond a bachelor's	18 years
Doctoral degree	18 years
"NA"	Omitted

TABLE D.2: IMPUTED EDUCATION YEARS FOR ACS DATA FROM BEFORE 2008

<i>educd</i> Code	Imputed Value
Nursery school to grade 4	2.5 years
Grade 5 or 6	5.5 years
Grade 7 or 8	7.5 years
High school graduate or GED	12 years
Some college, but less than 1 year	12 years
1 or more years of college credit, no degree	13 years
Associate's degree, type not specified	14 years
Master's degree	18 years
Professional degree beyond a bachelor's	18 years
Doctoral degree	18 years
"NA"	Omitted

## D.2 Survival Curves

### D.2.1 Adjustment of Survival Probability for Incarceration

In practice, since the data are discrete rather than continuous, we compute survival adjusted for incarceration, which is given in the model by  $\ell(t)(1 - m(t))$ , using the following equation:

$$\ell^{adj}(t) = \ell(t) - m(t-1) \cdot \ell(t-1), \quad (\text{D.1})$$

where we estimate  $m(t)$  for each age  $t$  using the average institutionalization probability by age. For the  $m(t, s)$  extension of the model, this equation becomes the following:

$$\ell^{adj}(t, s) = \ell(t) - m(t-1, s) \cdot \ell(t-1), \quad (\text{D.2})$$

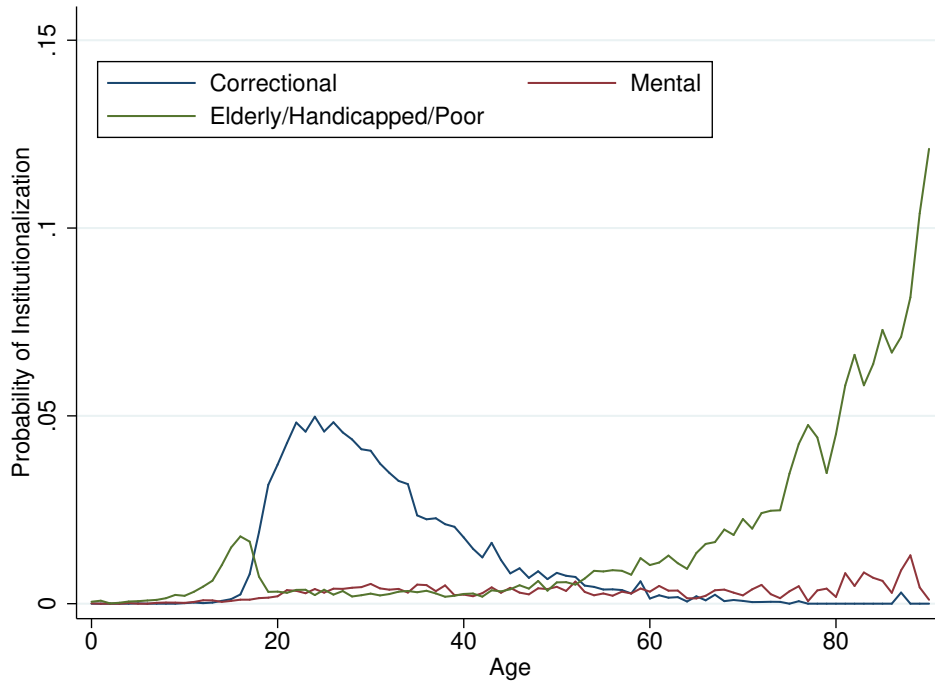
where we estimate  $m(t, s)$  for each age  $t$  and year of schooling  $s$  using the average institutionalization probability by age and education. For  $t < s$ , we impute  $m(t, s)$  as the school-independent  $m(t)$  probability, since  $m(t, s)$  for a fixed education level  $s$  cannot be observed in this range of age  $t$ .

As described in Section 4.3.5, we institute an age cutoff of 60 for the survival probability adjustment. Thus, for ages strictly above the cutoff, we assume that the incarceration rate  $m$  becomes zero, and we no longer perform the adjustments shown in equations (D.1) and (D.2). We determine the cutoff based on data from the 1980 ACS, which is the most recent survey to collect disaggregated institutionalization data. The graphs in Figure D.1 show age-based institutionalization rates by type amongst the Black and White populations in 1980.

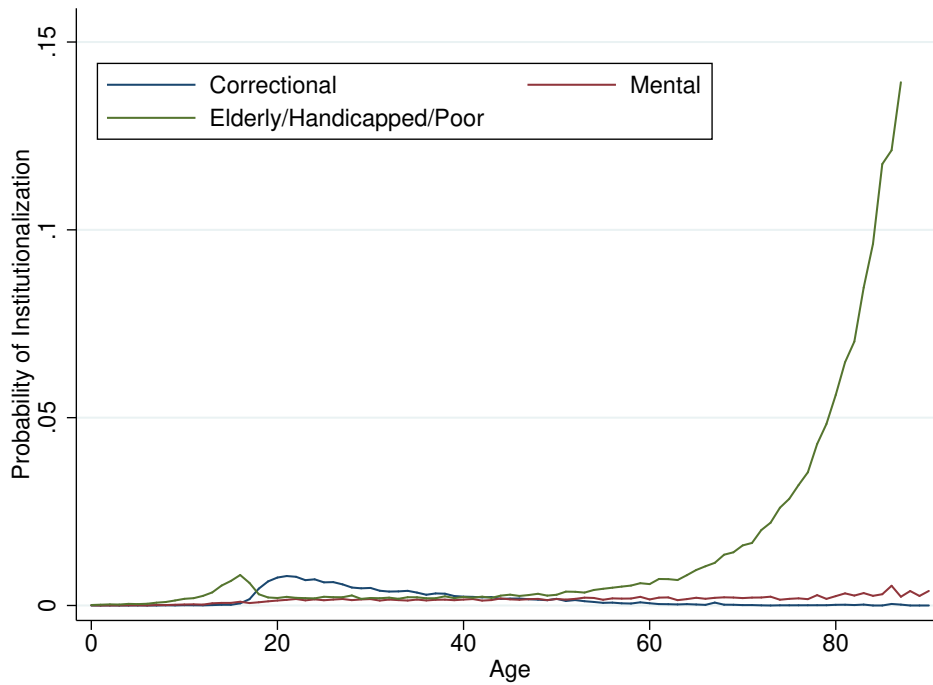


FIGURE D.1: INSTITUTIONALIZATION RATES BY TYPE FOR MEN IN 1980

(a) BLACK MEN



(b) WHITE MEN



Notes: These figures graph the institutionalization rates of Black and White men, by type of institution, against age, using the 1980 ACS data. For each institutionalization type, the corresponding curve consists of the average institutionalization probabilities for each age. For Panel B, institutionalization probabilities above 0.15 at the upper end of the age range are omitted from the graph.

## D.2.2 Formulation of Conditional Survival Curves

As noted in Section 4.1, the standard conditional survival curve (that is, corresponding to the mortality-only survival curve, unadjusted for incarceration) for age  $a$  is given by

$$\ell_a(t) = \frac{\ell(t)}{\ell(a)}, \quad (\text{D.3})$$

where  $\ell(a)$  is the probability of surviving to age  $a$ . To derive the conditional survival curve adjusted for incarceration, we use the same equation, except that the numerator becomes either  $\ell^{adj}(t)$  or  $\ell^{adj}(t, s)$  for the  $m(t)$  and  $m(t, s)$  cases, respectively, as in equations (D.1) and (D.2). Explicitly, for the  $m(t)$  case, the conditional, adjusted survival curve is given by

$$\ell_a^{adj}(t) = \frac{\ell^{adj}(t)}{\ell(a)}, \quad (\text{D.4})$$

while for the  $m(t, s)$  case, the conditional, adjusted survival curve is given by

$$\ell_a^{adj}(t, s) = \frac{\ell^{adj}(t, s)}{\ell(a)}. \quad (\text{D.5})$$

Thus, the conditional, adjusted survival curves are normalized by the probability of being alive at age  $a$ .

As highlighted in footnote 5, since the model conditions only on being alive at a given age, the setup assumes that individuals can continue schooling while incarcerated. An alternative but extreme assumption would be that individuals must stop schooling upon incarceration. To incorporate such an assumption into the analysis, the model would need to form a conditional expectation of the marginal benefit of schooling based on non-incarcerated individuals. Specifically, this approach would necessitate the following two alterations to the model's features. First, we would need to normalize the conditional, adjusted survival curves by the probability of being both alive and not incarcerated at age  $s_0$ , thus removing individuals who are incarcerated at age  $s_0$  from consideration. This normalizing probability would be drawn from the unconditional, adjusted survival curves that constitute the numerators in equations (D.4) and (D.5), rather than from the unconditional, unadjusted survival curve.

Problematically, in removing individuals who are incarcerated at the viewpoint age from the adjusted survival curve, this specification would render the conditional, adjusted survival curves fundamentally incomparable to the conditional, unadjusted survival curves. For the second alteration to the model, since we would now only be analyzing the remaining non-incarcerated individuals, we would need to condition the  $m(t)$  or  $m(t, s)$  function on the event of not being incarcerated at age  $s_0$ .

Lastly, in regard to equation (D.5), note that we compute the education-dependent survival curve in the numerator for fixed  $s$ , since the curve enters the  $\Delta$  formula, which we calculate separately for each  $s$ .

## D.3 Estimation Procedures

### D.3.1 Computation of Integrals

The estimation of several quantities in this study — including  $\Delta(s)$  and  $\frac{\partial\Delta}{\partial s}$  in all versions of the model and  $\frac{\partial^2\Delta}{\partial s^2}$  in the  $m(t, s)$  version of the model — requires the computation of integrals. To compute the integrals, we estimate a trapezoidal sum by partitioning the region under the integrand into one-year intervals, computing the trapezoidal area for each interval, and then summing all of the trapezoids to obtain the total area under the curve. As the life table data end at age 100 with a non-zero probability of survival, we impute the integrand for the  $\Delta(s)$  computation as zero at age 101 in order to close off the integral estimation. In practice, since the integrand attains such small values in this upper age interval, the imputation makes minimal difference to the estimation. For the  $\frac{\partial\Delta}{\partial s}$  and  $\frac{\partial^2\Delta}{\partial s^2}$  computations, such imputation is unnecessary because the age cutoff for incarceration adjustment (as described in Section 4.3.5 and Appendix D.2.1) means that the integrand values of the component integrals contained in the  $\frac{\partial\Delta}{\partial s}$  and  $\frac{\partial^2\Delta}{\partial s^2}$  formulas take on the value of zero for all ages above the cutoff.

### D.3.2 Estimation of $m(t, s)$ Derivatives

As shown in equations (4.19) and (4.21), estimates of the first and second derivatives of  $m(t, s)$  with respect to  $s$  are necessary to compute  $\frac{\partial\Delta}{\partial s}$  and  $\frac{\partial^2\Delta}{\partial s^2}$  for the  $m(t, s)$  version of the

model. To estimate  $\frac{\partial m}{\partial s}$  and  $\frac{\partial^2 m}{\partial s^2}$ , we begin by estimating a flexible, 4<sup>th</sup>-degree polynomial regression of institutionalization on age and education, done separately for Black and White men. Age and education terms are included in the regression for all powers, including all interactions between the terms. The samples used for the regressions are restricted to men within the age range between the starting year  $s_0$  and the age cutoff for incarceration adjustment. We then use the coefficient estimates to compute the first and second derivatives of  $m(t, s)$  with respect to education for Black and White men. For equation (4.19), note that  $\frac{\partial m}{\partial s}$  is evaluated at  $s_0$  years of schooling across all ages within the corresponding integral bounds. For equation (4.21), note that  $\frac{\partial m}{\partial s}$  is evaluated at  $s_0$  years of schooling and age  $s_0$ , while  $\frac{\partial^2 m}{\partial s^2}$  is evaluated at  $s_0$  years of schooling across all ages within the corresponding integral bounds.

### D.3.3 Estimation of Elasticity

Using each of the two elasticity estimation methods, we estimate the elasticity for all combinations of education level  $s$ , net interest rate  $r - g$ , education-on-incarceration gradient weight  $\alpha$ , and assumed White male elasticity  $\varepsilon_w$ . When estimating the elasticity, we first check the second-order condition for maximization of the lifetime income function. If the second-order condition is not satisfied (that is, if the second derivative is positive) for a given quadruple of education level, net interest rate,  $\alpha$ , and  $\varepsilon_w$ , we omit the corresponding elasticity computation from the overall estimation procedure. With our base specification of  $\varepsilon_w = 1$  and Elasticity Method 1, which uses equation (4.13), the second-order condition is violated in only seven of 832 estimation cases. These cases correspond to an education level of grade 9 with  $r - g \geq 0.03$  and  $\alpha = 1$  and an education level of grade 10 with  $r - g \geq 0.045$  and  $\alpha = 1$ . Thus, the second-order condition only fails at the  $\alpha$  weight upper bound and at interest rates that are greater than our preferred specification of  $r - g = 0.02$ . With  $\varepsilon_w = 1$  and Elasticity Method 2, which uses equation (4.14), the second-order condition is never violated. Across the complete set of combinations of education level, net interest rate,  $\alpha$ , and  $\varepsilon_w$ , the second order condition fails in 103 of 2,496 cases under Elasticity Method 1 and 55 of 2,496 cases under Elasticity Method 2.

### D.3.4 Estimation of Schooling Difference

In accordance with equation (4.5) and the overall model specification, the estimation procedure for the schooling difference involves estimating the Black-White gap for every combination of education level  $s$ , net interest rate  $r - g$ ,  $\alpha$ ,  $\varepsilon_w$ , elasticity estimation method, and schooling adjustment bound  $\underline{s}$ . In its entirety, the schooling difference estimation proceeds according to the following steps:

1. Estimate  $\Delta_b$  and  $\Delta_w$  for each year of schooling  $s$ , using a given net interest rate  $r - g$  and weight  $\alpha$ . For this estimation, we set  $s_0 = s$ . Therefore, the conditional, adjusted survival curve in the  $\Delta(s)$  equation takes the viewpoint of age  $s$ . In practice, while  $s_0 = s$  in the language of the mathematical model, we assume that the viewpoint age  $s_0$  is equal to  $s + 6$  when mapping the model to the data, as described in footnote 6.
2. Estimate  $\varepsilon_s$  with the procedure described in Section 4.3.4, using a given value for  $\varepsilon_w$  and either Elasticity Method 1 or Elasticity Method 2. For the  $m(t, s)$  version of the model, the component  $\frac{\partial \Delta}{\partial s}$  and  $\frac{\partial^2 \Delta}{\partial s^2}$  estimates again use the conditional, adjusted survival curve referenced to  $s_0 = s$  and the given net interest rate.
3. Use equation (4.4) to estimate the schooling difference at  $s$  based on the  $\Delta$  estimates from Step 1 and the elasticity estimate from Step 2.
4. Evaluate the maximum function in equation (4.5) for a chosen value of  $\underline{s}$ .
5. Repeat Steps 1-4 across all  $s$  and compute the final weighted average given by equation (4.5). Here, we omit cases in which  $s < \underline{s}$ , since the right-hand argument of the maximum function in equation (4.5),  $\underline{s} - s$ , which serves to provide a lower bound on the estimation, is positive for such cases. As the  $p_s$  weights for these cases are extremely small, the omission has little effect on the estimates.
6. Perform Steps 1-5 across all combinations of  $r - g$ ,  $\alpha$ ,  $\varepsilon_w$ , elasticity estimation method, and  $\underline{s}$ .

Across all of the individual estimation cases, the left-hand argument of the maximum function in equation (4.5) is the maximum in the majority of cases. In our base specification

of the model using  $\underline{s} = 7$ ,  $\alpha = 0$ ,  $\varepsilon_w = 1$ , and Elasticity Method 1, the right-hand argument maximizes the function in 13 of 130 cases. Across the complete set of combinations of education level, net interest rate,  $\alpha$ ,  $\varepsilon_w$ , elasticity estimation method, and  $\underline{s}$ , the right-hand argument applies in 1,720 of 9,360 cases. In general, the right-hand argument tends to kick-in for higher values of  $\alpha$  and for values of  $s$  near or equal to  $\underline{s}$ . As described in Section [4.4.1.2](#), the main results are robust to changes in  $\underline{s}$ .