ESSAYS ON EVALUATING A MATERNAL HEALTH VOUCHER PROGRAM IN RURAL UGANDA BY ITS IMPACT ON SERVICE DELIVERY AND QUALITY.

A Dissertation

Submitted to the National Graduate Institute for Policy Studies (GRIPS)

in partial Fulfillment of the Requirements for the Degree of

Ph.D. in International Development Studies

by

Nakawesi Claire Lwasa

April 2022

ABSTRACT

This dissertation evaluates a maternal voucher program implemented in the rural areas of Uganda. It aims to examine (i) the quantity-quality tradeoffs when decreasing the price of health service fees; (ii) the possible negative effects of financial incentives on Community Health Workers (CHWs)' activities that are not incentivized by the program; and (iii) program sustainability.

It consists of two main chapters. In the first chapter, I examine the program's sustainable impact on maternal care use, quality of the services, and health outcomes, both during the program implementation and after the program's withdrawal. I utilize two health-facility level panel data, as well as women's pregnancy-level panel data, and apply the difference-in-differences approach with health facility and women fixed effects respectively. The estimation results indicate that, during its implementation, the program enhances the utilization of maternal care (delivery, antenatal (4^{th} visit), and postnatal services), the perceived quality of these maternal health care services, and number of medical staff (doctors, nurses, and midwives). However, this impact on utilization and quality of services returns to pre-program levels after program withdrawal. In contrast, there is no distinct effect on maternal and child health outcomes. Except for the pro-poor effects on the number of medical doctors after the program's withdrawal, there is no conclusive evidence suggesting that the positive effects of the program on the usage and quality of maternal services are greater in poorer districts (in terms of per capita GDP) than in the other districts. These results support the implementation of comprehensive maternal voucher programs, but there is a need

to address concerns of child and maternal outcomes, and the sustainability of these positive effects beyond the program period.

In the second main chapter, I investigate if short-term financial incentives that are paid to CHWs affect their service delivery and work ethic. The difference-in-differences approach with CHW fixed effects is applied to CHW level panel data obtained from a survey in Uganda. The estimation results indicate that incentives increase service delivery and the working hours per week, and that these increases do not appear to be realized by crowding out the non-incentivized services, even though program CHWs decreased their working hours per week for other economic activities to increase service delivery. These effects are, however, unsustainable as the CHW effort levels return to pre-program levels once the incentives end. Tests for differential impact indicate no evidence that the effects of the incentives differed by gender, both during and after the program. However, the effects on household coverage and some services provided, were greater for CHWs who were selected through the community's popular vote than those who were not. Further, there were negative effects on the motivation of the poorer CHWs toward women counselling and health campaign services after the withdrawal of the incentives. These results illustrate that the efforts of CHWs are not sustained beyond the duration of the programs that support them, which brings to question what policy makers ought to do to sustain the performance of CHWs.

The findings of this dissertation contribute to the growing literature on the performance of voucher programs and provide timely public policy implications for improving health care service quality, utilization, and health outcomes.

DEDICATION

To my husband Julius Rumanyika and our beloved children, Kimberly Letitia Angelica Kihembo, Keith Louis Agaba, and Kristen Lucia Murungi, with love and gratitude.

ACKNOWLEDGEMENT

This journey has been worthwhile, thanks to support from various people. Most significantly, I acknowledge the support from my principal advisor, Prof. Yoko Kijima, who worked with me from the onset and patiently thought through my ideas as they evolved. Thank you so much Prof. Kijima for your time and energy, for your patience with me, for your kindness, tireless, and consistent support, and the positive feedback that has shaped my analytical, research and writing skills. I acknowledge my co-advisor, Prof. Yamauchi, with sincere gratitude for her constructive comments and suggestions. My gratitude also extends to Prof. Takahashi and Prof. Litschig, whose comments and suggestions at the beginning of writing this dissertation, helped me better structure my analyses.

My sincere thanks to the entire GCUBE faculty; Prof. Tetsushi Sonobe (former Program Director), Prof. Takahashi (current Program Director), Prof. Takagi, and Prof. Kijima for accepting me into the Ph.D. program, providing me with a fellowship which made my stay comfortable, and for the research fund, which supported the data collection from which my dissertation has been birthed. Special thanks are also due to all the GRIPS administrative staff, especially the GCUBE Course Coordinators (Awata Masayo and Mori Satomi).

I extend my gratitude to the Management and staff of the Ministry of Health in Uganda, Management and staff of Marie Stopes, Uganda — particularly the Director Programs, Mr. William Nyombi, and Program staff, Mrs. Juliet R. Atuhaire and Ms. Joyce Zalwango, for supporting my research work. Thanks to Mr. George Ssentumbwe and all

enumerators from Homeland Data Services for working through the hard times of the pandemic to deliver my data needs.

I am also thankful to the various students' bodies, my batchmates, and senior colleagues. I appreciate the academic and social engagements shared with various friends and colleagues, especially Dr. Mbithi A., Dr. Sorkpor C., Dr. Egami H., Dr. Curaporn C., Dr. Edem R., Dr. Owusu P., Dr. Njokwe G., Abebe Ambachew, Pascal Matthew Owor, Yin Min Htwe, Shahida Pervin Ruby. For my rich exposure to the Japanese culture and lifestyle, special thanks to my friends and their beloved families — Kaori Yamabayashi and Takae Afful. Thank you to the GRIPS Ugandan alumni and students — for your support. I also wish to appreciate my student friends from various universities in Japan, especially Caleb Kwame Sinclear (Tokyo Medical & Dental University) and Miller Kahuina (Tokyo University of Marine Science & Technology), for their encouragement and support despite the differences in our academic disciplines. I cannot forget the support received from the Embassy of Uganda in Tokyo — my thanks to you all.

I am highly indebted and thankful to my employer, Bank of Uganda, for accepting to support my ambitions and granting me study leave to pursue my studies at GRIPS. In the same spirit, I wish to thank my colleagues and friends, especially Mrs. Pelly R. Mugasi and Dr. Benedict Makanga, for their consistent support.

I am entirely grateful for the support given to me by my family (The Lwasa family, The Misanvu family, and The Bagerize family). To my beloved husband and children, it has been a long wait and I thank you so much for your patience, emotional support, and above all your prayers. Lastly, I thank the Almighty God for his provision and blessings.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

LIST OF ABBREVIATIONS

CAPI	Computer-Assisted Personal Interviews
CHW	Community Health Worker
DHO	District Health Office
DHT	District Health Team
DID	Difference in Differences
HMIS	Health Management Information System
IP	Implementing Partner
KES	Kenyan Shillings
MCH	Maternal and Child Health
MoH	Ministry of Health
PGT	Poverty Grading Tool
SDG	Sustainable Development Goal
TOT	Training of Trainers
UGX	Ugandan Shillings
URHVP	Uganda Reproductive Health Voucher Project
USD	United States Dollar
VHT	Village Health Team
VMA	Voucher Management Agency
VSP	Voucher Service Provider

CHAPTER 1

INTRODUCTION

1.1 BACKGROUND

Maternal and child health are extremely fundamental indicators of a country's socioeconomic development and quality of life as they characterize the population that is at a higher risk of ill health and death, and inform relevant strategies that are to be set to reduce these risks (De Janvry & Sadoulet, 2015). The Sustainable Development Goal (SDG) 3 on good health and wellbeing for all at all ages, among others, aims to decrease the global maternal mortality[1] ratio to less than 70 per 100,000 live births, and end the preventable deaths of newborns and children under 5 years of age by reducing neonatal mortality[2] to at least 12 per 1,000 live births, and do so by 2030 (WHO, 2016). To this date, women around the world continue to lose their lives due to pregnancy complications or childbirth, with most of these deaths (99%) occurring in developing countries — 64% percent occurring in the Sub-Saharan African region alone (WHO, 2018). The WHO emphasizes that the reduction of maternal and neonatal mortality crucially depends on women's access to quality care before, during, and after childbirth, and recommends that pregnant women take their early antenatal care in the first trimester as this can potentially reduce the risk of complications for women and newborns during and after child delivery. Unfortunately, in many developing

[1] "The probability of a woman dying while pregnant or within 42 days of termination of pregnancy, irrespective of the duration and the site of the pregnancy, from any cause related to or aggravated by the pregnancy or its management, but not from accidental or incidental causes" (WHO 2015).
[2] "The probability of an infant dying within the first month of life" (UDHS 2016).

countries, women face significant barriers to accessing formal healthcare services including limited health service availability, low service quality, socio-economic factors, and lack of education (Kawungezi et al., 2015; Yamauchi & Manang, 2019; Matsuoka et al., 2010; Riaz et al., 2015).

These barriers require a comprehensive approach that takes into account both the supply and demand side constraints to improve maternal health service utilization and health outcomes (Alfonso et al., 2013; Nguyen et al., 2012; Van de Poel et al., 2014). In recent years, policy makers have devised policies to address these persistent challenges by adopting financing initiatives, such as maternal vouchers, to reduce the cost barriers that impede women's access to healthcare so that they are encouraged to seek antenatal, delivery, and post-partum care from qualified service providers (Keya et al., 2018; Grépin et al., 2019). Available evidence indicates that vouchers can improve the usage of health services, improve health outcomes, reduce out-of-pocket costs for healthcare, and reduce the equity gap between the poor and the rich for access to health services (Nguyen et al., 2012; Ir, P. et al., 2010; Dennis et al., 2018; Pilasant et al., 2016; Bowser et al., 2016; Keya et al., 2018: Grépin et al., 2019). Most of this literature has evaluated the implementation of voucher programs in various countries to assess its impact on healthcare usage (antenatal visits, institutional deliveries, and postnatal care visits) among poor women and consequences on health outcomes. Unfortunately, little is known about the sustainable impact of these programs after they finish.

Furthermore, though the usage of services may increase, health outcomes may not necessarily be improved if the quality of service provision is not matched with the demand for the services. Evaluation of the Chiranjeevi Yojana voucher program in India showed that the voucher program did not improve service uptake if the facility was not ready, or if the quality of care was low (Mohanan M. et al., 2014). Furthermore, health facilities located in poorer communities are often understaffed and equipment can be poorly serviced or underutilized due to the lack of technical expertise (Obare et al., 2013). In many voucher programs, as was found in Rwanda, attention is paid to service outputs rather than quality, which ultimately affects the use of maternal health services at the facility (Skiles et al., 2013). Similarly, qualitative evidence from Kenya suggests that the free maternity services program overburdened the public health facilities and resulted in reduced motivation of the health workers thus reducing the quality of care (Tama E. et al., 2017). It is, therefore, worthwhile conducting a rigorous empirical study to establish the impact of these programs on the quality of service provision at the health facilities.

To explore the concerns about the sustainable usage of maternal care services and quality of service provision, I consider a comprehensive maternal voucher program, the Uganda Reproductive Health Voucher Program (URHVP), implemented in the rural areas of Uganda from 2015 to 2019. The program aimed to increase the demand and uptake of skilled maternal care services among the rural poor mothers, and to improve the provision of services at the health facilities through training, mentorship, support supervision, and service payments. Using annual administrative data (2014-2019) from Uganda's Ministry of Health and survey data that I collected from 140 health facilities and 275 rural poor mothers, I

examine the impact of the voucher program on the quantity of maternal care use, the quality of maternal care services, and health outcomes for mothers and children. Besides this, I also contribute to the growing body of literature by estimating the sustainable impact of the program after its completion.

Nyqvist et al. (2019) point to the importance of integrating Community Health Workers (CHWs) into health programs in order to improve the health care service utilization and health outcomes. The CHW strategy is popular in many developing countries, especially in sub-Saharan Africa and South Asia (Wagner, Z. et al., 2019). CHWs, who are traditionally volunteers, devote their time and effort to provide basic health care services, health related products, and to refer severe cases to the formal health care system, within the communities where they live. The roles of CHWs have been expanded and integrated in the formal health care system and their effectiveness is of paramount importance for the achievement of the SDGs (WHO, 2016).

Unfortunately, CHWs often fall short of their assigned duties because their motivation is low. This is due to the various challenges they face in the execution of their duties (Strachan et al., 2012). Therefore, an understanding of how to motivate CHWs is of great importance. In order to motivate their efforts and performance, there is a growing support in the literature for the use of entrepreneurial and financial incentives in CHW programs (Basinga et al., 2011; Miller et al., 2012; Finan et al. 2017; Nyqvist et al., 2019). Having said that, there are other studies that argue that pro-social CHWs perform better than financially motivated workers (Ashraf et al., 2014; Wagner, Z. et al., 2019). As to which kind

of motivation (financial or social) makes CHWs more effective in providing healthcare services, some concerns remain unanswered in the existing literature. Do the CHW services under the incentivized model crowd out other health services, which may negatively affect people with health problems other than the incentivized services, and does the mentality of CHWs toward volunteer work remain the same when the incentives are no longer provided? Both these possibilities may create negative unintended consequences. Since CHWs are socially motivated people and are usually engaged in various activities in the community (volunteer, charity, and political party activities), financial incentives to CHW services may crowd out such community activities. It is therefore crucial to understand if such an effect may be found as a result of the introduction of incentivized programs.

Taking the Uganda Reproductive Health Voucher Program (URHVP), I explore the possibility of these trade-offs. The CHWs in the program health facilities did not previously receive any financial rewards from the government for their services to the community. However, under this program, they were entitled to financial incentives: a transport allowance being the markup from the sale of each voucher (may be used for transport purposes or as additional income for those CHWs who could use a bicycle or walk), and a monthly bonus payment if they followed-up with their voucher clients to ensure the usage of the voucher. The financial incentives were aimed at boosting the efforts of CHWs in the delivery of the program services, while also reducing the challenges of low motivation and lack of transportation to make visits, as had been the case previously. Using my URHVP CHW survey of 140 health facilities and 272 CHWs, I provide, to the best of my knowledge, not only the first evidence if financial incentives increase CHW effort, but also if these

incentives crowd out the CHW's effort toward non-incentivized services and other community activities, and if the incentives affect the CHWs' work ethic.

In addition to the above contributions, since the URHVP included both the demand and supply side components, this study presents a complementary analysis of the effects of both components of the maternal voucher program.

1.2 MAIN FINDINGS

Unlike other maternal voucher-only programs elsewhere, the URHVP is a comprehensive voucher aimed at addressing both demand and supply-side constraints. The program not only provided free maternal care services to rural poor women, but also created supply-side financial incentives for health facilities and community health workers, to support the utilization and quality of maternal care services. Thus, this dissertation evaluates the URHVP's impact in three broad dimensions: i) on the demand for skilled maternal care services by poor rural mothers; ii) on the quality of service provision by providers; and iii) on the effort level of the CHWs as the agents between service provision and utilization. Despite the language used, there is no assertion of causality in the results obtained in this study.

In the first analysis, I find that the maternal voucher program enhanced women's utilization of maternal care (delivery, antenatal (4^{th} visit), and postnatal), the perceived quality of ANC, facility delivery care and PNC services, and the number of medical staff (doctors, nurses, and midwives) during the program implementation period. However, there

is no evidence of sustainable positive effects after the program was withdrawn. There is no evidence of improvements in maternal and child health outcomes and no findings that the positive effects of the treatment are greater in poorer districts than in better-off ones except for the number of medical doctors available after the program's withdrawal. Policy-wise, the results support the implementation of voucher programs for improving the utilization of maternal care services and quality of care despite there being an urgent need to solve issues of program sustainability and the improvement of health outcomes.

In the subsequent analysis, I find that the program increased the service delivery of CHWs in terms of the number of villages and households covered, and increased working hours per week. This increase is realized by decreasing the CHW's weekly hours spent on other economic activities. However, there is no evidence that the program decreased the CHWs' service provision of non-incentivized services, and the likelihood of participation in other community work. Yet, the estimated positive effects of the program on the service delivery of CHWs during the program period are unsustainable as they do not endure once the program is withdrawn. Policy-wise, financial incentives clearly are a powerful mechanism in enhancing the performance of CHWs, at least for the period that they last. Thus, the consideration of modest financial incentives for CHWs to facilitate their service delivery by reducing their transport burden may be helpful in sustaining CHW performance.

From the results, it is plausible that the increased usage of maternal care services is not only due to the cost subsidy enjoyed but possibly also due to the improved service provision and better motivated CHWs.

1.3 ORGANIZATION OF THE DISSERTATION

Chapter 2 examines if a comprehensive maternal voucher program has a sustainable impact on the use of maternal care, quality of services, and health outcomes, both in the program implementation and after withdrawal periods. I discuss these aspects based on the perceptions of health service providers and women, who are the beneficiaries of the maternal care services. I also present the existing literature and the research gap that I have attempted to fill. Finally, the description of the data, econometric model, and results are discussed.

Chapter 3 focuses on the short-term financial incentives paid to CHWs and examines if these incentives affect their service delivery and work ethic. I review the existing literature and contribute to it by focusing on the tradeoffs that come with the incentives. As in Chapter 2, I also describe the data, econometric model, and discuss the results.

Chapter 4 draws conclusions and presents the policy implications of the findings from the two analytical chapters.

CHAPTER 2

ESTIMATING THE IMPACT OF A MATERNAL HEALTH VOUCHER PROGRAM ON SUSTAINABLE UTILIZATION AND SERVICE QUALITY IN RURAL UGANDA.

2.1 INTRODUCTION

Maternal and newborn mortality are pressing policy challenges, particularly in developing regions, such as sub-Saharan Africa (SSA). In 2015, over 300,000 women lost their lives around the world due to pregnancy related causes resulting in a maternal mortality ratio (MMR) of 216 maternal deaths per 100,000 live births (UN Women, 2018). Almost all these deaths occurred in developing countries with 64% occurring in the SSA countries making it the region with the highest MMR (WHO, 2018). In the report that followed, in 2016, maternal mortality was second only to HIV/AIDS as the cause of death for women of reproductive age, especially in SSA (WHO, 2019). The reduction of maternal mortality crucially depends on women's access to quality care before, during, and after childbirth (WHO, 2019). The World Health Organization (WHO) recommends that pregnant women take their early antenatal care (ANC) in the first trimester (critical trimester) as this potentially reduces the risk of complications for women and their newborns during and after delivery. However, globally, many pregnant women (more than 40%) did not receive early ANC as of 2013 (WHO, 2018).

Maternal and child health constraints are common both in developed and developing countries, though the situation is dire in many developing countries where women face

significant barriers to accessing formal healthcare services, including limited health service availability, low service quality, socio-economic factors, and lack of education (Kawungezi et al., 2015; Yamauchi and Manang, 2019; Matsuoka et al., 2010; Riaz et al., 2015). These complex barriers require a comprehensive approach that considers both supply and demand side constraints to increase maternal health service utilization, level of service provision, and to improve health outcomes (Alfonso et al., 2013; Nguyen et al., 2012; Van de Poel et al., 2014).

So far, policymakers have addressed these challenges by adopting demand-side financial initiatives, such as vouchers and cash transfers to reduce the cost barriers and encourage women to seek antenatal, delivery, and post-partum care from qualified service providers (Keya et al., 2018; Grépin et al., 2019). These voucher schemes consist of partly or fully paid maternal care, where the holder is entitled to receive defined maternal services from an affiliated provider, whereas cash transfer schemes consist of a lumpsum transferred to the beneficiaries, usually with no guarantee that they will use the money to access maternal services (Agha, 2011a; Lagarde et al., 2009). While their mechanisms of financing may differ, their common aim is to reduce the financial barriers to access maternal care, and encourage service utilization to improve pregnancy outcomes, especially among the poor (Jehan et al., 2012). Maternal vouchers could increase the use of specific maternal care services; however, their effects are program- and context-specific and may depend on various social factors in the healthcare system (Alfonso et al., 2013; Nguyen et al., 2012; Van de Poel et al., 2014). While these programs are not all-inclusive (demand and supply side), most studies do not examine if the enhanced demand will endure after the program's conclusion,

and do not evaluate the programs' effect on maternal and child health outcomes (Hunter BM. et al. 2017).

The critical questions here are i) will maternal care for the poor cease when the price increases; and ii) will service levels decline since these governments cannot afford a perpetual subsidy due to fiscal constraints. To increase the demand for health products or services, Kremer et al. (2019) suggest that policymakers should avoid small health product price increases as they produce large take-up reductions. A randomized pricing experiment in Western Kenya showed that a USD 0 to USD 0.60 increase (100% to 90% subsidy) for insecticide-treated bed nets at prenatal clinics (still USD 0.15 below the market price) caused a 60% drop in demand by pregnant women (Cohen & Dupas, 2010). This suggests that once a subsidy is received, people equate the subsidized price to the market price (anchoring effect) and then become unwilling to pay the full market price. Similarly, in a Kenya "copay voucher" program, women were entitled to free ANC and postnatal care (PNC) services but had to pay USD 1.20 for delivery. This dramatically reduced voucher effectiveness (Grépin et al., 2019). Thus, cost-sharing programs are less effective than free distribution programs due to their discontinuous impact on the demand for health services. Further, Fischer et al. (2019) conducted a field experiment in northern Uganda where some curative health products were distributed door-to-door, either free by an NGO, or for sale by a for-profit company. It was found that for all the health products, the subsequent purchase rates lowered after a free distribution.

Meanwhile, Dupas (2014) used data from a 2-stage randomization procedure to distinguish between the anchoring effect and learning effect. She dismissed the fears of reference dependence as temporary subsidies can increase short-term adoption among subsidy recipients and their neighbors. This consequently increases the willingness to pay or causes long-term demand for health products through the learning effect when the information remains salient. Given this premise, I examine the sustained effects of a maternal voucher program in Uganda on the usage of maternal care services not only during program implementation, but also after program withdrawal.

As noted earlier, low service quality significantly inhibits women from accessing healthcare services. Thus, research is required on the effect that maternal voucher programs have on service quality. Even if the program enhances demand, health outcomes may not be improved due to various issues including congestion, long waiting time, frequent stockouts of medicines and sundries, and staff work overload. In terms of facility delivery, research shows that a public conditional cash transfer program in India enhanced institutional deliveries, but did not improve mother and infant health. This is likely because of reduced service quality (Lim et al., 2010; Powell-Jackson et al., 2015). Although some studies find that fee abolitions and maternal vouchers may have reduced service quality, rigorous evaluation is limited (Grepin et al., 2019). It is therefore on this premise, that I attempt to estimate, the impact of a voucher program on service quality and health outcomes.

I considered a maternal voucher program, the Uganda Reproductive Health Voucher Program (URHVP), implemented in the rural areas of Uganda between 2015 to 2019, as my

case study. This comprehensive program had two components: i) to empower women, thereby increasing the demand and uptake of ANC, safe delivery, and PNC through maternal vouchers for rural poor mothers; ii) to improve the health facilities by strengthening the institutional capacity to deliver reproductive health services through training, mentorship, support supervision, and service payments. Similar to programs implemented in other countries, this program targeted poor and vulnerable pregnant women in rural areas who were incapable of meeting the costs for maternal care services. The program was implemented by a specialized sexual and reproductive healthcare organization, Marie Stopes International – Uganda, who acted as the Voucher Management Agency (VMA). Eligible Voucher Service Providers (VSPs) — public or private — and Village Health Teams (VHTs) (*hence forth referred to as Community Health Workers (CHWs)*) were assessed by the VMA and respective district health officials for their capacity to deliver maternal care services. The CHWs visited the women's homes and used a customized poverty grading tool (face-to-face survey) to collect data on demographics, pregnancy status, prior births, household welfare, and assets. This information was then used to assess the pregnant woman's and the household's poverty level. Only those women who obtained a score of 0-12 out of 21 points were considered poor and eligible to purchase the voucher. The voucher cost the eligible women UGX 4,000 (USD 1.09) and entitled the holder to four antenatal care visits, safe and skilled delivery at the health facility (including a caesarean section if necessary), emergency transportation to a referral facility, and postnatal care (within six weeks after birth) for the

mother and newborn at accredited healthcare providers (public and non-public[3]). The URHVP entitled the women to a very subsidized package as indicated in Table 2A1 (Panel A), which is more generous than voucher programs implemented in other countries, such as the Reproductive Health Voucher Program in Kenya where a similar voucher cost the beneficiary KES 200 (USD 2.70) (Dennis ML. et al., 2018). All eligible candidates purchased the voucher; thus, there was no self-selection to buy or not buy the voucher, so long as the evaluation process by the CHWs was not manipulated. Accredited health facilities received regular training, mentorship, support supervision, and payments for the services provided.

As a part of this study, 140 health facilities and 275 mothers from 30 districts were surveyed, and I collected the panel datasets which cover periods before, during, and after the program. Combined with the annual administrative data from the Ministry of Health (MoH), I estimated the program's impact on both the quantity and quality of maternal care services, and maternal and child health outcomes by using the Difference-in-Differences (DID) with fixed effects estimation approach.

The estimation results show that the voucher program increased the utilization (antenatal 4th visit, facility delivery and postnatal care), and the perceived quality of maternal care services during the program implementation period. However, the program's positive impact did not endure after the program's withdrawal, though the number of mid-wives and

[3] Public facilities refer to health facilities which are government owned and fully funded by the government, while non-public facilities refer to the health facilities which are owned and funded by non-government organizations, religious organizations, individual entrepreneurs, or self-employed practitioners (MoH, National Health Facility Master List 2018).

delivery beds were sustained even after the program ended. Furthermore, I do not find evidence that the program improved maternal and child health outcomes. Lastly, since the impact of the program on maternal health service usage and quality may be larger in districts with households who are poorer, and initially worse-off in health outcomes, I also estimated heterogenous treatment effects based on per capita GDP of the district. There is evidence of an increased number of medical doctors in the health facility after the program's withdrawal, which suggests that the program did have pro-poor effects on this outcome. There is, however, no evidence that the program's positive effects on the utilization and quality of services during the program period were any larger in better-off districts than the poorer ones. This suggests that the program's positive effects on usage and service quality were not concentrated in the wealthier districts, and that the poorer districts also benefited from this program.

The rest of the chapter is organized as follows. Section 2.2 reviews the literature, while Section 2.3 presents the context on maternal health services in Uganda and a description of the voucher program. Section 2.4 details the data used for the analyses, followed by Section 2.5, which covers the empirical models used. Section 2.6 presents the estimation results and discussion, and Section 2.7 concludes this chapter.

2.2 LITERATURE REVIEW

The principle of subsidizing the use of health services, popularly known as demand-side financing, has been in existence for over 15 years now; either through direct cost sharing with users, or through an income effect provided to households, to motivate health seeking behavior (Hunter BM. et al., 2017). There are many forms of user-focused financing in the health sector, namely, vouchers for maternity care services. They are designed to encourage attention toward maternal and newborn health by reducing the cost of maternity care services at the point of use. The vouchers entitle holders to specific maternal care services that must be sought from a pre-approved set of public or private service providers.

As reviewed in the systematic study by Hunter BM. et al. (2017), many studies have reported that maternal vouchers improve the demand for antenatal care, skilled attendants at birth, birth in healthcare facilities, and postnatal care (for example, Nguyen et al., 2012; Van de Poel et al., 2014; Mohanan M. et al., 2014: Alfonso et al., 2015). However, these studies only focus on the program implementation period, which is typically two to three years after program commencement. Therefore, conclusions cannot be drawn about their long-term effects on usage and quality of service care. Additionally, these studies do not examine the impact the programs have on maternal and child outcomes. Some recent rigorous studies related to the impact of maternal voucher programs in developing countries do provide evidence on the impact of maternal vouchers on the usage of maternal care services, and like the earlier studies, the effects are program and context specific and may depend on various social factors and conditions within the healthcare system (Keya et al. 2018; Grépin et al. 2019; Dennis ML. et al. 2018).

In Bangladesh, the maternal healthcare voucher scheme introduced in 2006 was implemented by the government and mostly enrolled public health facilities. It targeted the pregnant poor rural women in select sub-districts (upazilas). Eligible women, identified by the field workers, were entitled to free services (three antenatal care visits, facility delivery (normal and cesarean), management of complications, emergency referral, and postnatal care services); free medicine for complications and delivery; and a cash stipend for transportation. Mothers were also entitled to a conditional cash transfer of USD 29 and a gift box, if they delivered with a skilled birth attendant at home or at the facility. The scheme also reimbursed the participating facilities with funds that were divided proportionately between the designated staff and a facility management fund. This scheme was therefore a combination of supply side incentives for health care providers, and demand side cash transfers and subsidies for the service users.

Keya et al. (2018) analyzed the effect of this voucher program on increasing deliveries with skilled care either at home or in facilities, and reducing the access disparity between the rich and poor pregnant women with respect to the use of the facilities. Their findings, using the difference-in-differences model on cross-sectional data, suggest that the voucher program significantly increased the use of public health facilities (13.9 percentage points) and seeking delivery complication management care at the facility (13.2 percentage points), though an insignificant effect was found in the increase of facility-based deliveries. A sub-group assessment of the 5 well-functioning facilities indicated that the number of facility deliveries increased by 5.3 percentage points. A quintile-based analysis of all the facilities showed that facility deliveries increased more than three times for lower quintile households, but only

about two times for the control sites. While this study informs us of the positive effects the program had on the usage of delivery services by poor women, it does not provide any additional findings on the use of other maternal care services, quality of care, and health outcomes.

Grépin et al. (2019) conducted a randomized control experiment in Kenya's Vihinga District to test if maternal voucher programs, cash transfers (to lessen transport costs burdens), reminder SMSes, and free healthcare policy were effective at enhancing the use of maternal care services among the rural poor pregnant women. Women assigned to the "full voucher arm" were entitled to free ANC, delivery, and PNC services, while those assigned to the "copay voucher arm" were only entitled to free antenatal and postnatal services, but had to pay USD 1.20 at a time for a facility delivery. Women assigned to the conditional cash transfer were paid transport costs for up to four antenatal care visits (USD 3 per visit), a facility delivery (USD 6), and three postnatal care visits (USD 3 per visit) upon presentation at the health facility, while those assigned to the unconditional arm received the same amounts before each of their scheduled visits to the health facility. Some women were also selected to receive weekly SMSes reminding them of forthcoming antenatal and postnatal care visits, and encouraging them to deliver at the health facilities. Some women received a "plain" text version reminding them to attend their scheduled visits and the facility delivery, while others received the "contextualized" text version, which was a modification of the plain type plus baby health as a salient feature. Lastly, the free care policy by the government exempted all pregnant women from paying for all maternity services in all public facilities across the country. Using the linear probability model, the authors found that full vouchers

and conditional cash transfers were extremely effective in improving the rates of facility deliveries while the copayment vouchers only had a very small and insignificant effect. They also found no evidence that the unconditional cash transfers, text messages, and the free care policy had a significant impact on increasing the rates of facility deliveries. None of the treatments had any effects on the demand for antenatal and postnatal services. This is likely because the attendance rates for these services were already relatively high in the sample. Thus, this study also informs us of the positive and varying effects of the various approaches to increasing facility deliveries. However, conclusions on antenatal and postnatal services are only based on the endline survey, which provides biased estimates of these outcomes. Also, the effects on self-reported quality were found to be weak due to the limited availability of data.

On another account in Kenya, the Reproductive Health Voucher Program was implemented in selected counties to provide impoverished women with subsidized access to public and private sector care. Vouchers were sold to eligible women at KES 200 (USD 2.70), thus entitling the eligible holders to four ANC visits, facility delivery (normal or caesarean), and a PNC visit. The program also aimed at expanding provider choices and improvements in the quality of care at both public and private lower-levels and referral facilities. During its implementation, the Kenyan government introduced the free care policy exempting all pregnant women from paying for all maternity services including ANC, delivery, and PNC services in all public facilities throughout the country. Using the difference-in-differences approach on data from cross-sectional surveys, Dennis et al. (2018) evaluated the reproductive voucher program for evidence of its short term and long-term effects (2006-

2016) on the utilization of maternal care services, and if any observed effects of the voucher program persisted after the introduction of free maternity services (2013). Taking period1 for the pre-voucher period, period2 for the full voucher implementation, and period3 for the period after free public maternity services were introduced, they show that there was an increase in the utilization of maternity services in both the voucher and non-voucher sub counties — 4+ ANC visits increased from 59.4% in period1 to 62.7% in period2, and a further moderate increase in period3; facility deliveries increased from 50% in period1 to 83.2% in period2 and to 86.7% in period3; and postnatal care increased from nearly 60% in period1 and period2 to 82.1% in period3. The results also indicated that the use of private sector facilities increased substantially between periods1 and 2 for all types of maternal care services in the voucher counties. However, this declined between period2 and period3 across both voucher and non-voucher counties, even though it remained significantly high in the voucher districts. My study comes close to this study by estimating the impact of the program on the use of maternal care services even after its withdrawal, while this study extends to the period when free public maternity services were introduced. Although this study demonstrates the impact of the initiatives on the utilization of all maternal care services, it lacks the much-needed analysis on the quality of care and health outcomes over the three periods.

My study contributes to the maternal care literature on the impact of voucher programs (for maternity care services) on the use, service provision of maternal care services, and maternal and child health outcomes by providing evidence from the administrative data of the MoH, Uganda, and data from my own health facility and women surveys during the

program implementation, and after the withdrawal of the program. I also add to the research on the heterogenous effects of voucher programs by establishing whether the treatment effects of the URHVP are greater in wealthier areas than in poorer areas by using the secondary sub-national or district level GDP data collected using the Enhanced Light Intensity Model.

2.3 CONTEXT

This section provides a brief background of the trends of maternal care services in Uganda and the Uganda Reproductive Health Voucher Program (URHVP).

2.3.1 MATERNAL HEALTH SERVICES AND HEALTH FACILITIES IN UGANDA

Over the years, Uganda has not only made progress in improving the provision and use of maternal care services, but also reduced the rates of maternal and infant mortality. However, in comparison to other peer countries in Africa, her progress still lags behind. Nevertheless, health care is a component of the social transformation aspiration in the country's Vision 2040, with the following strategies being proposed for the sector: "empower households and communities to promote healthy lifestyles and practices, utilize a preventive health care system as a more sustainable approach, enhance the nutrition of children and women of reproductive age, and shift from a public centered health delivery system to a public-private-partnership" (Uganda Vision 2040, 2017).[4] Through the National Development Plans —

[4] Uganda Vision 2040, 2017 accessed online on June 16, 2021, from http://library.health.go.ug/publications/leadership-and-governance/uganda-vision-2040

NDP1 (2010/11-2014/15), NDPII (2015/16-2019/20), and NDPIII (2020/21-2024/25)[5] — and in line with the SDGs, the government of Uganda has made various efforts at improving child and maternal health outcomes. Specifically, as a part of the sectoral plans and strategies, the Health Sector Development Plans (HSDP 2010/11-2014/15 & 2015/16-2019/20)[6] aim to achieve the nationally set targets by improving the government's oversight over the health sector, expanding private and donor partnerships with the government, increasing channels of service delivery, including health information, financing, products and technologies, workforce, and infrastructure (Obare et al., 2016; UDHS, 2016).

Trends indicate improvements in the usage of maternal care services in Uganda[7]. Recent findings indicate that 97% of the women, who were aged 15-49 with a livebirth at least 5 years prior to the last Uganda Demographic Health Survey (UDHS) in 2016, received antenatal care (ANC) from a qualified provider during their most recent pregnancy. However, only 29% of women initiated their ANC visits during the first trimester of pregnancy, while 60% completed at least four ANC visits. Although this surpasses the 45% target set in the

[5] All NDP Plans accessed online on June 16, 2021, from http://www.npa.go.ug/development-plans/national-development-plan-ndp/

[6] HSDP accessed online on June 16, 2021, from http://library.health.go.ug/publications/work-plans/health-sector-development-plan-201516-201920

[7] In the last 16 years there has been an improvement in the percentage of women receiving antenatal care from a qualified provider at least once during their most recent birth, from 90% in 2000-01 to 93% in 2006; 95% in 2011 to 97% in 2016. The percentage of women who took least four ANC visits rose from 42% in 2000- 01 to 47% in 2006; 48% in 2011 to 60% 2016. Deliveries from recognized health providers grew from 37% in 2000-01 to 42% in 2006; 57% in 2011 to 73% in 2016. Home deliveries fell from 62% in 2000-01 to 58% in 2016; 42% in 2011 to 25% in 2016. The presence of a qualified service provider during childbirth has also improved over the past 16 years, from 37% in 2000-01 to 42% in 2006; 58% in 2011 to 74% in 2016.

.

Health Sector Development plan, it still falls short of the ideal of universal coverage. National statistics on institutional deliveries indicate that more than 73% of livebirths in the 5 years preceding the survey were delivered in a health facility and with a qualified birth attendant (74%). Home deliveries decreased by more than half from 62% in 2000-01 to 25% in 2016, following the expansion of the national health facility network. For postnatal care, only 54% and 56% of women and newborns, respectively received a postnatal check within 2 days of the delivery for the most recent births in the 2 years preceding the survey. While these numbers are encouraging, rural women are still less likely to seek skilled maternal care services than urban women —7% vs 18% for ANC; 70% vs 88% for deliveries in the facility; and 51% vs 67% for PNC — largely due to financial constraints (UDHS 2016).

It is further reported that the neonatal mortality declined from 33 deaths per 1,000 live births in 2000-01 to 27 deaths per 1,000 live births in 2006 and has remained stagnant; whereas infant mortality reduced from 88 deaths per 1,000 live births in 2000-01 to 43 deaths per 1,000 live births in 2016 (UDHS 2016). Maternal mortality rates also declined from 524 deaths per 100,000 live births in 2000-01 to 368 deaths per 100,000 live births in 2016. Health-facility-based maternal mortality declined over time from 168 deaths per 100,000 live births in 2012-13 to 148 deaths per 100,000 live births in 2016-17 (UBOS Statistical Abstract, 2019).

Improving maternal care services by bringing health facilities closer to the public has been a long-standing objective of the government. Since 1999, governments have emphasized improving access to health facilities by expanding the national healthcare infrastructure. Progress on this has been slow owing to the tight and unsteady budget

allocations to the health sector. For example, government allocation to health in the year 2017/18 was 6.4 percent of the total national budget, down from the 8.9 percent spent in 2016/17 (UBOS Statistical Abstract, 2019). It moved again to 7.2 percent in 2018/19 and to 5.1 percent in the 2020/21 budget, down from 7.9 percent in 2019/20 (UBOS Statistical Abstract, 2020). The number of functional healthcare facilities increased by 533 additional health facilities, from 6,404 facilities in 2016/17 to 6,937 in 2017/18 (MoH HF Master list, 2018). The health facilities are classified into seven categories depending on the services they provide and the geographical areas they are intended to serve[8]. They are classified as Health Centre Level one (HC I) to Health Centre Level four (HC IV), General Hospitals, Regional Referral Hospitals, and National Referral Hospitals. Health facilities are owned by the Government, Private and Not-For-Profit (PNFP), Private for Profit (PFP), and by the community (MoH HF Master List, 2018). In this study I refer to public facilities as government owned and non-public facilities as PNFP (including mission/religious entities), PFP, or community owned. I also define higher-level facilities as any facility that is HCIII and above, while lower-level facilities are HCI, HCII, and clinics.

[8] Health Center I (HCI) located at the village provides community based preventative and promotive health services for 1,000 individuals; Health Center II (HCII) located at the parish provides preventive, promotive and outpatient curative, outreach care and emergency services for 5,000 individuals; Health Center III (HCIII) located at the sub-county provides preventive, promotive and outpatient curative, maternity, inpatient and laboratory services for 20,000 individuals; Health Center IV (HCIV) located at the county provides services offered by HCIII, emergency surgery and blood transfusion services for 100,000 individuals; General hospital located at the district provides services offered by HCIV, other general services, training, consultation and research for 500,000 individuals; Referral hospitals provide services offered by general hospital and other specialized services for 1 million individuals :Regional referral hospitals all services offered by referral hospital and other specialist services and higher level surgical services for 2 million individuals; and national referral hospital provides comprehensive specialist services, training and research for 10 million individuals (MoH Master List, 2018).

Despite these achievements, there is still a need to encourage the sustainable use and improved quality of maternal care services in Uganda to achieve better child and maternal outcomes, especially among the rural poor communities, where access and quality are still of great concern.

2.3.2 THE UGANDA REPRODUCTIVE HEALTH VOUCHER PROGRAM (URHVP)

The URHVP aimed to encourage the use of skilled maternal healthcare services among rural poor women living in disadvantaged areas. These services would be available during pregnancy, delivery, and the postnatal period. The project also aimed at improving the quality of services provided at health facilities. The four-year project started in September 2015, though participating facilities started receiving the benefits at different points in time and ended at all facilities in August 2019[9]. Expected to support 142,438 safe deliveries, the project was approved in 2014 and funded by World Bank (World Bank, 2014). The project was implemented by a specialized sexual and reproductive healthcare organization, Marie Stopes International – Uganda, who acted as the Voucher Management Agency (VMA). URHVP sought to improve women's empowerment, increase the demand and uptake of antenatal care (ANC), safe delivery, postnatal care (PNC), and strengthen institutional capacity to deliver reproductive health services — all part of an effort towards Uganda's social transformation and attainment of Sustainable Development Goals 3.1 and 3.2, to

[9] Without much detail, I was advised that the program started earlier with facilities in Western Uganda. Therefore, I am not yet aware of any exogenous reasons which caused this difference in program rollout at participating facilities.

reduce global maternal mortality to less than 70 per 100 000 live births, and to reduce neonatal mortality to at least as low as 12 per 1000 live births by 2030, respectively.

Broadly, the project had two main aims. First, to provide the rural poor women who purchased a voucher with free access to: (1) four ANC visits including screening, advice on birth preparedness, and family planning; (2) safe and skilled delivery (normal, assisted, or caesarean); (3) emergency transport if a woman needed to be referred to a higher-level facility; and (4) PNC including newborn care and family planning within six weeks after childbirth. The second aim was to build national capacity in the health sector through training, mentorship, and supportive supervision of the participating healthcare providers in partnership with the respective health teams.

The project was implemented in 25 districts; 12 districts in the western region of Uganda (Mbarara, Kabale, Kanungu, Ntungamo, Kiruhura, Sheema, Buhweju, Mitooma, Ibanda, Isingiro, Bushenyi, and Rubirizi) and 13 districts in the eastern region of Uganda (Jinja, Bugiri, Kamuli, Buyende, Kaliro, Luuka, Mayuge, Iganga, Namutumba, Kibuuku, Tororo, Namayingo, and Busia) as indicated in Figure 2.1. The district's eligibility for voucher distribution was based on the district-level health indicator proxied by the percentage of children delivered at institutions (public facilities and private-not-for-profit facilities) out of all the children born in a given period. Therefore, districts deemed eligible had district health indicators that were lower than the national average of 44%. The primary project beneficiaries were the poor and vulnerable pregnant mothers who could not afford out of pocket expenses for their maternal healthcare services. Eligible women had to reside in the

catchment areas of the health facilities, that is, within 5kms or 2 hours travel time from the health facility. The secondary beneficiaries of the project were the contracted service providers and surrounding communities.

Through the public procurement provisions and the respective District Health Officers (DHO), health service providers (both public and non-public health) were publicly encouraged to submit their Expression of Interest (EOI) to participate in the program. Using criteria developed by the MoH, the facilities were assessed on several aspects including their registration with the government; physical state of the facility; qualifications of the health workers; quality of the tools and equipment used; and their interest in the program. The assessments also included site visits by the District Health Teams (DHT) and the VMA, to confirm if the facilities meet the clinical requirements to be able to provide safe maternity services. Providers who scored 65% and above in this assessment were considered eligible for the program. If more than one health facility was within a 5km radius, the facility with the highest score was selected. For functional referral networks, facilities were clustered to the nearest referral facility (which may be in another district), that is, each cluster had one referral facility and four to six facilities. The health facilities that qualified (hence forth Voucher Service Providers (VSPs)) were oriented and trained, after which, two CHWs were attached for support.

According to the program, the CHWs selected the eligible pregnant women. They visited these women with a customized poverty grading tool (face-to-face survey) and collected data on demographics, pregnancy status, prior births, household welfare, and assets,

among other factors. This information was used to assess the women's poverty level. Only pregnant women who scored 0-12 out of 21 points were considered eligible for the voucher purchase. The VMA conducted regular audits to ensure that the CHWs did not manipulate the poverty grading tool in favor of better-off mothers. Once confirmed as eligible, participants were informed of the program's benefits, and upon receiving their family's consent, were invited to purchase the voucher. All the eligible women purchased the voucher. The vouchers were sold to the CHWs at UGX 2,700 (USD 0.73)[10] per voucher and the CHWs in turn sold the vouchers to the eligible women at UGX 4,000 (USD 1.09). The markup of UGX 1,300 (USD 0.35) was used as the CHW's transport allowance (or additional income for those CHWs who could use a bicycle or walk). CHWs also received a monthly bonus payment of UGX 30,000 (USD 8.16) if they followed-up with their clients to ensure voucher usage and achieved a redemption rate of over 80% per month. The pregnant women who purchased the voucher were given a hardcopy voucher which had a unique ID number as indicated in Figure 2.2. They were also given information of the participating VSPs within their catchment area and were told to present the voucher at the health facilities when seeking any of the services under the voucher package.

The VSPs received full reimbursements within 20 working days following the submission of their monthly claims to the VMA in accordance with the standardized costs in Table 2A1 (Panel A). The VMA used the MarieTXT SMS system to monitor the voucher distribution

[10] Using exchange rate for September 01, 2015: USD1 = UGX3,676.36, retrieved from https://www.bou.or.ug/bou/collateral/interbank_forms/2015/Sep/Major_01Sep2015.html

and sales by the CHW, and service provision by the VSP. VSPs received annual clinical audits, mentorship, and supportive supervision to assess the quality of care provided and the adherence to the service guidelines issued by the VMA and DHT. The VSP audits ensured that the facilities maintained sufficient medical supplies and sundries to meet the anticipated demand, as well as maintaining the high quality of care and standards.

2.3.3 RESEARCH QUESTIONS

The URHVP aimed to encourage the use of skilled maternal healthcare services among rural poor women living in disadvantaged areas during pregnancy, delivery and postnatal period. The project also aimed at improving the quality of services provided at the health facilities. Thus, this study's objectives are to investigate how a maternal voucher program (URHVP) can affect the sustainable usage of maternal care, service quality, and maternal and child health outcomes.

Firstly, I test the prediction that URHVP improved women's usage of maternal care services at the health facilities, during program implementation and after it ends. This is measured by the number of ANC visits, deliveries and PNC visits. My concern is examining if the usage of maternal care services at the health facility decreased when the program ended. I analyze this by measuring the program's impact on the usage of all maternal care services after the program withdrawal. Although the usage of maternal care services may increase, this may still not lead to the desired and sustainable health outcomes if the quality of service provision is not matched to the demand for services. So, in my second analysis, I examine the prediction that URHVP improved the quality of services provided at the health facilities

during and after the program's implementation. This is measured by the service provider's self-assessment and the women's assessment of the service providers. Thirdly, I expect the URHVP to improve maternal and child health outcomes. This is measured by the incidence of birth complications, the birth weight of infants, and pregnancy outcomes. Lastly, as the URHVP aimed to provide free access to maternal care for the rural poor women and to improve health service quality, I examine the claim that the program's impact on the usage and quality of maternal health services may be larger in districts with poorer households — as they are initially worse-off in most health outcomes.

Therefore, I expect to answer the following research questions: a) Did the URHVP improve the usage of maternal care services? This is measured by the number of ANC visits, deliveries, and PNC visits at the health facilities during and after the program's implementation. b) Did the URHVP improve service quality at the health facilities during and after program's implementation, measured by the service provider's self-assessment and women's assessment of the service providers? c) Did the URHVP improve maternal and child health outcomes, which are measured by the incidence of birth complications, the birth weight of infants, and pregnancy outcomes? d) Were the treatment effects greater in poorer areas than in wealthier areas?

Why URHVP? Unlike other maternal voucher-only programs elsewhere, that focused largely on cost reduction for beneficiaries, this program aimed in addition, to support health service providers and agents to increase service availability and quality in poor rural settings.

The URHVP, therefore, presents an opportunity to provide new insights into the usage and provision of maternal care services.

2.4 Data and Descriptive Statistics

Data for my analyses came from the health facility administrative annual data on Maternal and Child Health (MCH) from the Ministry of Health's HMIS Form 105 for 2014 to 2019, as well as health-facility level and women-level surveys (referred to as the URHVP Health Facility survey and the URHVP Women Survey), both of which were conducted between January and March 2020 by the authors. Table 2A2 (Panel A and B) gives detailed information on the samples for this study.

2.4.1 MoH HMIS Data

I used the health facility annual data on maternal and childcare service provision from the MoH HMIS 105 forms for 2014, 2015, 2016, 2017, 2018, and 2019, for 140 health facilities – totaling to 840 observations. This is administrative data compiled by the MoH from the monthly reports submitted by all the health facilities. With the permission of the MoH, I obtained the data from Ministry's Statistics Office. For the purposes of this study, I restricted myself to the Antenatal, Maternity, and Postanal sub-sections. This annual dataset provided the total number of women receiving maternal care services at the respective facilities in the defined periods. It provides data on the usage of maternal care services at the health facilities. This dataset does not contain detailed information about the health facilities, such as service quality and capacity. Thus, I had to collect my own data for these factors.

2.4.2 URHVP HEALTH FACILITY SURVEY AND URHVP WOMEN SURVEY

I conducted the URHVP Health survey in 140 health facilities across 30 districts over 3 regions (Central, East, and West); 12 districts did not receive the program, while 18 districts did. My survey data constituted 51 program health facilities (supported by URHVP) and 89 non-program facilities (not supported by URHVP). The non-program facilities were selected from both program and non-program districts (23 and 66, respectively).

As explained in Section 2.3.2, the program covered 25 districts. First, I selected 18 out of the 25 URHVP project districts (9 in the western region and 9 in the eastern region) based on the district health indicator (if it was less than or equal to the national average of 44% in 2014)[11]. For the non-program districts, I selected districts (3 in the Western region, 3 in the Eastern region, and 6 in the Central region) that were neighboring the program districts and had comparable maternal health conditions to those of the program districts. This was achieved by setting a threshold index level of 21% and above (percentage of children delivered at institutions out of all the children born in a given time). Second, from the URHVP project documents, I found that there are 110 program health facilities in the 18 selected districts. As indicated in Section 2.3.2, the program facilities were evaluated to assess their capacity to provide safe maternity services, and only those that obtained a score of 65% and above were considered eligible to participate in the program. There is no known exogenous factor that affected the selection processes. To ensure statistical relevance, given

[11] Dropped were outliers (Luuka at 18% in the eastern region) and those that exceeded the national average (Bushenyi (57%), Isingiro (48%), Mbarara (65%) from the western region and Busia (54%), Iganga (50%), Jinja (65%) from the eastern region.

a limited budget to allow data collection from 140 health facilities, I used stratified random sampling for the selection of sample health facilities. After dividing the program health facilities into strata; where strata were district, subcounty, ownership, and levels (HCI, II, III, IV or clinics), I randomly chose sample program health facilities from each stratum, by taking every second program health facility. Third, after gathering the list of all health facilities in the sampled program and non-program districts from the Ministry of Health's Statistics Office, (from a total of 1,154 non-program health facilities), I also divided them into strata (district, subcounty, ownership, and levels); I then sampled one out of every 13 facilities for the non-program facilities, from each stratum. This resulted in the selection of 66 non-program facilities from the non-program districts and 23 non-program facilities from the program districts. The MoH and DHO granted me the permission to conduct interviews with the heads of the facilities and with the participating mothers at their respective health facilities.

Data were collected using structured questionnaires by CAPI (Survey Solutions). The URHVP Health facility survey was conducted in English, while the URHVP Women survey was in three local languages — Luganda for the Central region, Lusoga for the Eastern region, and Runyakitara for the Western region. Pilot and pre-test exercises were conducted during the enumerator training to confirm the accuracy of the tools and the respondent's confidence in the recall data before actual data collection commenced. Both surveys sought retrospective data from before the program (period 2011-2014), during the program (period 2015-2019), and after the program's withdrawal (6-month period between 2019 and 2020). As these periods were distinctive, the respondents could easily remember the data requested,

while some health facility data were retrieved from the health facility records. I however, acknowledge the possibility of recall error in the responses to retrospective questions. A team of 11 experienced and trained enumerators collected the data under close supervision.

Each health facility had one respondent who was either a facility in-charge or deputy in-charge. If the facility in-charge or deputy was not available, or had not worked at the facility for at least 5 years, any other official who met the criteria was used for the interview. A total of 140 respondents were interviewed regarding their age, education, years of experience, function at the facility, as well as their residence details. They were also interviewed about specifics, such as, the level of the facility, the nearest or associated referral facility, the operating hours, issues with service provision, and support from any maternal programs since 2011. Data for three periods (pre-program, program implementation, after program withdrawal) focused on the facility's source of funding, capacity, patients, staff, equipment, services offered, perception of service issues, as well as CHWs and their roles. Interviews were also conducted on the URHVP program (regarding selection into the program, the URHVP treatment, CHW roles, and perceptions of service provision under the URHVP). I created a 3-period panel dataset from this information.

Sampling of the women's survey was done using the snowball technique. In each facility, I asked CHWs to list the women who had experienced more than two pregnancies, with at least one in the pre-program period, and at least one during the program implementation period. Thus, the women respondents from the program facilities had to be URHVP beneficiaries for at least one pregnancy. I selected 2 women at each health facility.

We, thus, interviewed 275 respondents[12] (97 from the program facilities, and 178 from the non-program facilities) on their age, marital status, religion, education, residence details, biological children, and dependents. We also interviewed them about their recent pregnancies in the last 10 years; methods used to delay or avoid pregnancy; and their perceptions on the provision of maternal health services by the CHWs, medical staff, and the health facilities. Data for three periods (pre-program, program implementation, and after program withdrawal) focused on their occupation, assets, and access to financial services. Respondents from the program facilities were also interviewed on the URHVP program (regarding their selection for the program and perceptions about the program). A 3-period panel dataset was constructed from this information. However, study results from the URHVP women survey are based on the 2-period mother-level panel.

2.4.3 DESCRIPTIVE STATISTICS

Here I discuss the detailed summary statistics for the health facilities and women. I categorized the health facilities as program and non-program facilities. Table 2.1 indicates that 64.3% of the facilities are publicly owned with a larger concentration of 70.8% belonging to the non-program facilities. Further, 67.9% of the facilities in the sample are of the higher-level status, with 80.4% of them belonging to the program facilities. This means that half the program facilities were public and most of them are of the higher-level status. Other factors,

[12] The survey targeted 2 women from each health facility (2*140=280) however, we had 3 less respondents from the program facilities (Kibuku HCIV, Taoky Medical Clinic HCII, Shuuku HCIV), 3 less respondents from the non-program facilities (Burambira HCII, Rushaka HCII, Bugobero HCIV), and 1 extra respondent from a non-program facility (Budadiri HCIV).

such as the distance from the health facility to the DHO, distance to the referral facilities, common mode of transportation, travel time to the health facility (in minutes), and checking if Community Health Workers are attached to facility, were similar between the two study groups.

In Table 2.2, I show the descriptive results comparing the program and non-program facilities on the usage of maternal care services, bed and staff numbers, and the quality of services in the pre-program, program implementation (post1), and post-program withdrawal period (post2). I also show the changes that occurred over time in the three periods for program and non-program health facilities. In columns 6 and 7, I use the symbol '[a]' to indicate the significant differences over time ($p<0.05$) in the outcome variables in each study group for the periods labeled pre (before URHVP) and post1; in columns 10 and 11, I use the symbol '[b]' to indicate the significant differences over time in the outcome variables in each study group for periods labeled pre and post2; I similarly use the symbol '[c]' to indicate the significant differences over time in the outcome variables in each study group for the periods labeled post1 and post2. The four variables in the first segment include the average number of women who received maternal care services (ANC, facility delivery, and PNC) per year in each health facility. These variables come from the MoH HMIS Data. The remaining variables are from the URHVP Health Facility Survey, whose data was collected for three periods – the last 6 months in 2019-20 (after the program's withdrawal; post2), 2015-19 (program implementation period; post1), and 2011-14 (pre-program period). The measures

in the second segment are the bed and staff numbers[13] at the facility and are regarded as the objective measures of service quality. The third segment comprises the service quality measures with the first two variables focusing on the attention that inpatient pregnant women got from the medical staff[14], while the next three variables are the subjective measures of service quality[15].

It is important to note that the pre-program outcome variables are not different between the program and non-program facilities (except for the number of times an inpatient is served per day). The significant differences between the program and non-program facilities are largely noted in the program implementation period, and in some instances, in the program withdrawal period, as follows:

- In the 1st segment during the program period, there was a greater usage of ANC, delivery, and PNC services at the program facilities in contrast to the non-program facilities;
- In the 2nd segment, the program facilities had a higher number of beds (inpatient maternity beds and delivery beds) and medical staff (medical doctors and midwives); and

[13] In (*insert period*), what was the total number of (*inpatient medical general beds, inpatient maternity beds, delivery beds*) at this facility? Number of (*medical doctors, nurses, midwives*) filled in *(insert period*).
[14] In the (*insert period*), what was the average number of times per day an inpatient was served by medical personnel; and in the (*insert period*), what was the average time (in minutes) that an inpatient waited to be served by medical personnel at a time?
[15] In the (*insert period*), how would you rate the quality of (*insert service*) in terms of quality? 01 Very poor, 02 Poor, 03 Fair, 04 Good and 05 Very Good. I defined a dummy taking one if response is "Very Good" and zero if otherwise (Good, Fair, Poor and Very Poor).

- In the 3rd segment, the program facilities had better quality ANC, delivery, and PNC services.

For differences seen over time that are statistically different from zero, there is evidence that on average, the usage of ANC (4th visit) and facility delivery services increased over time at the program facilities alone. By contrast, the usage of PNC services increased at the program and non-program facilities during the program implementation and after the program's withdrawal. Temporally, there was an overall improvement in the bed and staff numbers at the facilities, though these differences were small and only strong for delivery beds at the program facilities. For service quality measures, I observed that the subjective level of quality improved, especially at the program facilities during the program's implementation, while the quality of ANC services decreased once the program was withdrawn. In Table 2A3, I present more statistics for the health facilities on their waiting facilities, other medical staff, and the CHWs attached to the health facility. I note that the study groups are comparable across all periods except for the increased number of medical staff (theater assistants, clinical officers, and anesthetists) at the program facilities during the program period. There are no significant differences seen over time. Lastly, in Table 2A4, I provide descriptive statistics for the health facility respondents. They show that the respondents from both study groups were largely comparable, except for the larger share of female respondents in the non-program facilities (66.3% vs 43.1%), and the larger share of married respondents from the program facilities (86.3% vs 71.9%).

The women respondents were categorized by the program and non-program facilities, and the statistics in Table 2.3 show no major differences between these two categories. On average, the women were 32 years old; 82.9% were Christian; they had 4 biological children and had been pregnant at least 3 times in the last 10 years (2010-2020); 93.1% had attended school and 52.7% had completed at least the primary level. The level of literacy among women is generally low (as shown by their proficiency in reading, writing, and speaking English). Radio is the most accessible source of information with 78.5% of the women listening to radio at least once a week. On average, women hardly watch television and use the internet, especially those from the program facilities. I also noted differences in the marital status and distance from their homes to the nearest health facility. While the higher share of women in the program facilities are married (96.9%) when compared to the non-program facilities (91.6%), the distance to the facility is closer for those in the non-program facilities (2.5km, taking 25.6 minutes to walk) than those in the program facilities (3.9km, taking 35.6 minutes to walk). 97.5% of the women are aware of the methods used to avoid or delay pregnancy, with the bigger concentration being the non-program beneficiaries. In Table 2A5, I show that across all the periods, more program women's spouses were engaged in subsistence farming, and that the program women possess lesser assets in comparison to the non-program women. For differences observed over time, I find that there is a general increase in asset possession and access to financial services across both study groups.

I also present the pregnancy outcome variables for the period 2010-2019. Table 2.4 shows the characteristics of the program and non-program women on the usage of maternal care services, quality of services, and maternal and child outcomes in the pre-program and

program implementation periods. I also show the changes over time in the two periods for the program and non-program women. Owing to the insignificant number of pregnancies (14 cases) after the program's withdrawal, I cannot explore the program's impact after withdrawal. The first set of variables shows the usage of skilled maternal care services. For each pregnancy, women were asked where they went for ANC, how many times they took ANC, where they went for the delivery, and where they went for PNC (6-8 weeks) after birth. To measure the impact of the program on the usage of ANC, delivery, and PNC, I defined four dummy variables. Each dummy variable equals 1 if a mother chose to take the maternal care service (ANC, delivery, and PNC) at any formal health facility, be it government, private, or mission/religious facility. The cases where a mother chose a Traditional Birth Attendant (TBA), or did not go anywhere, take the value of 0. The second set of variables are quality assessment measures for the maternal services received for each pregnancy[16], as provided by the CHWs, health facility, and medical staff, if the woman actually had contact with them. The last set of variables relate to the maternal and child health outcomes of each pregnancy. In the women's survey data for pregnancies over the last 10 years, women were asked if they experienced any complications during childbirth.[17] The survey also asked if the babies were weighed and, if yes, the weight in kilograms. Although we asked about the baby weight at birth, some (9.49%) of the answers obtained were 5kgs and over, which is likely to be the baby's weight a few weeks or months after birth. As indicated, the average infant

[16] How would you rate the services provided by (*insert VHT or health facility or medical staff*)? 01 Very poor, 02 Poor, 03 Fair, 04 Good and 05 Very Good. I defined a dummy taking one if response is "Very Good" and zero if otherwise (Good, Fair, Poor and Very Poor).

[17] My data indicates that pregnancies with complications were 9.54%, among which 4.96% had assisted and cesarean section delivery.

birthweight is not low at the program and non-program facilities, even in the pre-program period (3.6 kgs).[18] Lastly, for the pregnancy outcomes, women were asked whether they had a live birth or not for each pregnancy.[19]

It is important to note that the pre-program outcome variables are not different between the program and non-program women (except for the likelihood of facility deliveries, and the baby being weighed at birth). In the program implementation period, significant differences are noted between the program and non-program women, as follows:

- In the 1st segment, program women are more likely to take facility deliveries and PNC services. In comparison with the pre-program period, there was a significant increase over time in the incidence of facility deliveries for both program and non-program women during the program implementation period. There was an increased incidence of women from the program facilities taking four or more ANC visits and PNC services for each pregnancy.
- In the 2nd segment, program women had a higher likelihood to receive quality services from the CHWs, health facilities, and the medical staff during pregnancy. This also increased significantly over time.
- In the 3rd segment, more program women had their babies weighed at birth and experienced complications during birth (even when I consider women

[18] From my data, the share of infants with low birthweight is only about 8%.
[19] Did you have a live birth (*insert pregnancy order*)? 01 Yes, 02 No (stillbirth), 03 No (miscarriage), 04 No (died after birth but before 1 month after birth), 05 No (still pregnant). My data indicates that the pregnancies that resulted into a livebirth were 95.22%, stillbirths were 1.23%, miscarriages were 2.33% and neonatal deaths were 1.10%.

who experienced complications at birth and had assisted and cesarean section delivery). I am, however, not certain why pregnancies with complications during delivery increased at the program facilities. There are no differences over time in the child and maternal health outcomes at all facilities, except for the increase in the incidence of baby weight being measured at birth at the program facilities. This may be due to increased facility delivery.

- Lastly, for the pregnancy outcomes my results indicate no significant overtime differences in livebirths, stillbirths, miscarriages, and neonatal deaths in both study groups.

2.5 EMPIRICAL MODEL

The descriptive statistics showed that, there was an overall increased usage of skilled maternal care services, number of beds and number of medical staff, and the provision of quality services at the program facilities during the program period. However, some of the effects did not endure after the program's withdrawal. For the service users, there was an increased incidence of facility deliveries, an uptake of PNC services, positive reviews on the receipt of substantially better quality services from the CHWs, health facilities, and medical staff, an increased incidence of complications during delivery, and an increase in the measuring of baby birth-weight among the program women.

My goal is to examine the impact of the URHVP on i) the usage of maternal health care services; ii) the quality of maternal care services provided at health facilities; and iii) maternal and child health outcomes. In this section, I conduct regression analyses to achieve

this goal. Having controlled for unobserved health facility heterogeneities and year fixed effects to control for common shocks, as well as district-specific linear time trends to account for the possibility that the services evolved differently in the program and control facilities, the identification strategy states that the treatment status of the program and error terms in the service provision model are not correlated.

Table 2A2 (Panel C) provides detailed information on the data descriptions for this study. On the utilization of maternal care services, I examine i) the MoH annual data for the average number of women per year who took the 1st ANC, 4th ANC, Delivery, and PNC at the health facility; and ii) the pregnancy-level survey data for the likelihood of a woman taking ANC, 4 or more ANC visits, delivery, and PNC services at a facility for each pregnancy. On the service quality at the health facility, I use i) the health facility survey data from quality self-assessments of the ANC, Delivery, and PNC services at the facility that were reported to be very good, the number of times an inpatient received service in a day, and the waiting time for inpatients before they received service; ii) the health facility survey data of the number of health facility beds (general beds, maternity, and delivery beds) and staff (doctors, nurses, and midwives). This is because better equipped and staffed facilities tend to lead to better health care; and iii) the women pregnancy-level survey data as service users reporting if the quality of maternal care provided by the CHWs, health facility, and medical staff was very good. Lastly, I use the women pregnancy-level survey data to assess the maternal and child health outcomes based on the incidence of complications during childbirth, infant birth weight, and pregnancy outcomes.

Although the URHVP implementation started in 2015, only 25.49% of health facilities received the program in 2015, while 68.63% received the program in 2016, and 5.88% received the program in 2017. The program ended at all facilities in 2019. To calculate the usage of maternal care services from the MoH facility-level data, I used the annual raw data for each facility from 2014 to 2019, taking the pre-program period as 2014, program implementation period as 2015 to 2018 (post1), and post-program withdrawal period as 2019 (post2). For the health facility survey, the data was collected in average terms for the respective periods, specified as — before URHVP, during URHVP, and after URHVP ended. Therefore, for the quality measures from the health facility survey, I use the data as it was collected for the pre-program (2011-2014), program implementation (post1; 2015-2019), and post-program withdrawal (post2; 6 months in 2019-20).

The women survey data collection was similar to the quality measures from the health facility survey. Further, I obtained women pregnancy-level data for the last 10 years (2010 to 2019) for both program and non-program beneficiaries. The yearly pregnancies are classified into two periods — pre-program pregnancies (2010 to 2014); and program period pregnancies (2015 to 2019). As stated above, there were too few pregnancies after the program withdrawal to researching any impacts in the post program withdrawal period.

First, using the MoH annual data and the health facility survey data, I employed the difference-in-differences approach with health facility level fixed effects. I used the following estimation models to investigate the average effects of the URHVP on the usage of maternal health care services and service quality during the program's implementation

(post1) and after the implementation (post2); where 2.1a is used for MoH annual data and 2.1b is used for health facility survey data:

(2.1a) $Y_{jdt} = \delta_0 + \delta_1 P_{jdt}^1 + \delta_2 P_{jdt}^2 + \gamma_{dt} + \phi_j + T_t + \varepsilon_{jdt}$,

where Y_{jdt} indicates the natural log of the average number of women who took the 1st ANC, 4th ANC, Delivery, and PNC checkup at facility *j* in district *d* in year *t*. P_{jdt}^1 is the program indicator variable equaling 1 for facilities under the program in year $t \in$ (2015-2018) and P_{jdt}^2 is the program indicator variable equaling 1 for facilities under the program in year $t = 2019$. γ_{dt} are district specific linear time trends. δ_1 and δ_2 are the coefficients of interest for impacts during the program implementation post1 (2015-2018), and after the program withdrawal post2 (2019), respectively. ϕ_j are health facility fixed effects which control for unobserved variables that differ from one health facility to another, but remain constant over time, while T_t are year fixed effects controlling for unobserved variables that are the same across facilities but differ over time. ε_{jdt} is the error term clustered at district-level. Uganda follows a decentralized health model, implying that DHOs are tasked with the planning, management, and service delivery (Henriksson et al., 2019). Hence, I clustered standard errors at the district level to correct the standard errors for potential correlations of temporal outcomes for the health facilities within a district.

(2.1b) $Y_{jdt} = \delta_0 + \delta_1 Post1_t + \delta_2 Post2_t + \delta_3 (P_j \times Post1_t) + \delta_4 (P_j \times Post2_t) + \gamma_{dt} + \varphi_j + \varepsilon_{jdt}$,

where Y_{jdt} indicates the following outcomes of interest at health facility *j* in district *d* in period *t*: a) the natural log of the average number of women who received the 1st ANC, 4th ANC, Delivery at facility *j*, and PNC checkup per year in the period *t*; b) an indicator variable that equals 1 if the respondents considered the ANC, delivery, and PNC services provided by the health facility *j* at period *t* to be very good, and 0 otherwise; c) the number of times per day that an inpatient was served by medical personnel, and the time (in minutes) that an inpatient waited to be served each time at health facility *j* in period *t*; and d) the number of beds (medical general, maternity, and delivery) and medical staff (doctors, nurses, and midwives) at health facility *j* in period *t*. $Post1_t$ and $Post2_t$ are indicator variables that equal 1 for the program-implementation period and post-program withdrawal period, respectively. P_j is the program indicator variable equaling 1 if a health facility *j* was supported by the program. $\delta_3$and δ_4 are the coefficients of interest for impacts during the program (post1), and after the program withdrawal (post2), respectively.

As participating facilities started to receive the program at different points (25.49% in 2015, 68.63% in 2016 and 5.88% in 2017), one may be concerned about the influence of some exogenous factors which may have caused this rollout.[20] In an attempt to exploit this, using equation 2.1a, I compared facilities served in 2015 (treated group) with those that started in 2016 (control group) and also, I compare those that started in 2015 and 2016

[20] Without much detail, I was advised that the program started earlier with facilities in Western Uganda. As I could not verify why, I checked whether proximity to the District Health Office (as a proxy for distance to the district town) mattered. I find that, irrespective of the region, the distance to district town is closer for facilities that received the program earlier. The average distance (km) from the health facility to District Health Office is 11.73km, 17.13km and 19.33km for facilities that started to receive the program in 2015, 2016 and 2017 respectively. Therefore, it is not realistic to assume that program rollout was done randomly.

(treated group) with those that started in 2017 (control group). Unfortunately, in both cases, the effect is not detected probably because of the small number (only 51 in total) of program facilities (results not shown for brevity). One might also be concerned about the possibility of the program effects being contaminated by two aspects. Firstly, some women who are not in the catchment area of the program facilities take services at the facilities. Although this is possible, I was assured by the VMA that every woman's voucher was tied to a specific facility. Thus, the eligible women had to reside in the catchment areas of the contracted health facilities, i.e., within 5 kms or 2 hours travel time from the health facility. Thus, this may not be possible as the voucher is facility specific. Secondly, the program trained CHWs and medical staff at the facilities moved to other facilities. For example, a CHW moves within the same district and the performance of health facilities then improves in the non-program facilities. CHWs were constrained to sell vouchers only within their designated / catchment areas. I also checked how commonly health facility staff moved from program to non-program facilities during the program implementation period and found that 32 respondents moved from one facility to another. However, only 2 out of the 32 moved from program facilities to non-program facilities. Based on this, I believe that the possibility of contamination is very negligible as staff movements from program to non-program facilities were not common.

For the second regression, I used pregnancy level data for the last10 years (2010-2019) to also examine the average effects of URHVP on the use of maternal care services, quality of maternal services, and the maternal and child health outcomes based on the difference-in-differences women fixed effects model:

(2.2) $Y_{ijdt} = \beta_0 + \beta_1 P_{jdt} + \mathrm{W}_{ijdt} + \gamma_{dt} + \mu_i + T_t + \varepsilon_{ijdt},$

where Y_{ijdt} indicates the following outcomes of interest for woman *i's* pregnancy at health facility *j* in district *d* in year *t*: a) an indicator variable that equals 1 if a woman *i* took ANC, delivery, and PNC services for a pregnancy at year *t* at a health facility *j*, and 0 otherwise; b) an indicator variable that equals 1 if a woman *i* evaluated the maternal care services provided by the CHWs, health facility, and medical staff for a pregnancy at year *t*, as very good, and 0 otherwise; and c) an indicator variable that equals 1 if a woman *i* had complications at birth for a pregnancy at year *t*, and 0 otherwise; d) birth weight of the baby in kilograms at year t and; e) an indicator variable that equals 1 if a woman *i*'s pregnancy resulted in a livebirth, stillbirth, miscarriage, or neonatal death at year *t*. P_{jdt} is the indicator variable that equals 1 for facilities under the URHVP program in year $t \in$ (2015-2019). β_1 is the coefficient of interest for the impact of the program on a woman. W_{ijdt} is a set of time-variant women characteristics such as the woman's age at each pregnancy; child characteristics (sex, birth order); woman and spouse's primary source of income; and possession of household assets like agricultural land, livestock, poultry, and durables (TV, radio, motorcycle, bicycle, house). γ_{jdt} are district specific linear time trends. μ_i are the women fixed effects. ε_{ijdt} is the error term clustered at district level. Similarly, I cluster at the district level to correct the standard errors for any potential correlation of outcomes over time for mothers within a district, including those mothers with repeated pregnancies.

The identification strategy is based on the common trends assumption. However, my data from the MoH has one pre-program period (2014). Although the URHVP started in

2015, only 25.49% of the health facilities received the program that year — an indication that the actual implementation of the program was not active in the starting year. Therefore, I test if the pre-program trend is the same for program and non-program facilities by using the MoH administrative data for 2014 and 2015. I use the estimation model below:

(2.3a) $Y_{jdt} = \delta_0 + \delta_1 P_{jdt} + \gamma_{dt} + \varphi_j + T_t + \varepsilon_{jdt}$,

where δ_1 is the coefficient of interest that, if found to be insignificant, implies that there is no difference between the program and non-program facilities in the pre-treatment trend. The results in Table 2A8 (Panel A) show that the pre-treatment trend is the same for the program and non-program facilities. Similarly, using the women pregnancy level data for 2010 to 2014, I used the estimation model below to test if the pre-program trend is the same for the program and non-program women.

(2.3b) $Y_{ijdt} = \beta_0 + \beta_1(P_i \times Post_t) + W_{ijdt} + \gamma_{dt} + \mu_i + T_t + \varepsilon_{ijdt}$,

where $Post_t$ is the indicator variable that equals 1 for pregnancies in year t = 2014 and 0 for pregnancies in year $t \in$ (2010-2013); P_i is the program indicator variable equaling 1 if a woman i was supported by the program; and β_1 is the coefficient of interest, which indicates no difference between program and non-program women in the pre-treatment period as shown in Table 2A8 (Panel B). Although I attempt to verify the parallel-trend assumption, no claim to causality is made.

To ameliorate any concerns about false rejections due to use of many outcome variables, I employ the Romano-Wolf correction method, which controls the family-wise error rate (FWER). I include adjusted p-values in regression results that belong to a family of outcomes (Clarke et al 2019).

2.6 RESULTS AND DISCUSSION

2.6.1 ESTIMATION RESULTS

2.6.1.1 IMPACT OF URHVP ON THE USAGE OF MATERNAL CARE SERVICES

In Table 2.5, I present the estimated impact of the URHVP on the usage of maternal care services (ANC, delivery, and PNC). Although I present the results with and without district specific linear time trends, the main discussion focuses on results with the influence of district specific linear time trends. Table 2.5 (Panel A) shows the estimation results for the usage of maternal care services from the MoH administrative data. From the results, there is evidence of increased usage of ANC, facility delivery and PNC services at the program facilities during program implementation by 38.1% (ANC4), 76.6% (delivery), and 56.3% (PNC), as indicated in row program_t1, columns 6 to 8. In the program withdrawal period, there is no difference in the usage of all maternal care services between the program and non-program facilities (indicated in row program_t2, columns 5 to 8). Thus. the effect on the usage of maternal care services is not as strong in the post-program withdrawal period. However, since there is no significant difference in the coefficients of Program_t1 and Program_t2, this shows that the impact of the program on ANC, delivery, and PNC services remains sustainable even after the program's withdrawal. If the value of skilled maternal care

services is salient, mothers will continue to seek these services despite the lack of a subsidy. This is the justification for providing a subsidy. These results may further suggest that the demand for maternal care services is not price elastic, and that the learning effects from the subsidy program are large enough to overcome the cost burdens.

Table 2.5 (Panel B) shows the estimation results from the URHVP Women survey. It indicates that although there is an increased likelihood of women using delivery services (by 15.2 percentage points) at the program health facilities, there is no difference in likelihood of usage of ANC and PNC services between program and non-program beneficiaries (indicated in row program_t, columns 5 to 8). It is likely that the attendance rates for ANC and PNC services were already relatively high in the sample. The results in Panels A and B are inconsistent. This is probably because the sample of the women survey is not nationally representative, as explained in the data section. Particularly, I relied on the CHWs to select women with at least two pregnancies in the last 10 years. This means that those who had had only one pregnancy, and were most likely less experienced with being pregnant, were not included in the Women survey data. Thus, the positive effects found in Panel A (but not in Panel B) might be because such women took ANCs, delivery, and PNC services in the program facility. The positive on facility delivery found in Panel B means that even among experienced mothers, relative to the nationally representative sample and the unobserved mothers' heterogeneity, the program effectively enhanced facility delivery. Furthermore, looking at the costs of maternal care services indicated in Table 2A1, women were interviewed on which maternal care services they paid for, for each pregnancy and how much they paid in Uganda Shillings, if at all they paid anything. Panel A shows the URHVP

standardized costs incurred by the facilities if they provided a specific service, whereas Panel B reports the actual out-of-pocket costs incurred by the women for the respective maternal care services. The out-of-pocket expenses in Panel B are lower than the program standardized costs in Panel A. This is possibly an indication that the non-program beneficiaries may be able to access these services at public facilities (constituting 70.8% of the sample for non-program facilities), even in the absence of subsidies for maternal care services, especially if the value of the skilled maternal care services is salient.

Taken together, these results indicate that the URHVP enhanced the utilization of maternal care services during the program period. However, given the selection limitations, the results for the women analysis should be interpreted with caution. Although the results using the women's data are unable to show the impact of the program after its withdrawal, the health facility results show that the usage of maternal care services was sustained (albeit in small proportions). This is because even in the absence of a subsidy, skilled medical care is salient in the mother's mind. Unfortunately, though, the health providers perceptions (Table 2A6) about the URHVP indicate that the program increased congestion and overloaded the working staff at the program facilities, albeit in small proportions. Thus, in the next subsection, the impact of the program on the service quality is examined.

2.6.1.2 Impact of URHVP on the quality of maternal care service provision

In Table 2.6 (Panel A), I show the estimation results for the impact of the program on the service quality during the program's implementation (post1) and after the program's withdrawal (post2). First, the results of the subjective measure of quality evaluated by the

health facilities indicated that in comparison with the pre-program period, there was an increase in the likelihood that very good delivery and PNC services were provided (at the program and non-program facilities) during the program implementation period (indicated in row post1). However, this increased more among the program facilities — by 32.6 percentage points and 30.7 percentage points, respectively (indicated in row programxpost1 and columns 7 and 8 respectively). The program also enhanced the quality of ANC services at the program facilities by 29.8 percentage points (indicated in row programxpost1 and columns 6). In the period following the program's withdrawal (post2), there was no difference observed in the quality of service provision for all the maternal services between program and non-program facilities (indicated in row program x post2, columns 6-8). Since the program facilities only took advantage of the program to improve their quality of ANC and PNC services during the program implementation period (row program x post1=program x post2, columns 6 and 8), it is concluded that this was an indication of unsustainable service quality after the withdrawal of the program.

Secondly, I use the women pregnancy level data for the period 2010-2019. These data provided service quality measures evaluated by the users for CHWs, health facility, and the medical staff. Table 2.6 (Panel B) shows the estimation results. Although there are improvements in the quality of services provided by the medical staff during the program period — by 29.7 percentage points (indicated in row program_t, and column 6), there was no evidence found that suggested improvements in the quality of the service delivery by

CHWs and the health facilities (indicated in row program_t, and columns 4 and 5)[21]. In addition to their initial evaluation and acceptance into the program, health facilities were subjected to annual clinical audits, mentorship, and support supervision to assess the quality of care and adherence to the service guidelines issued by the VMA and the DHT. If found short, the VSP was disqualified from the program.[22] It is therefore, puzzling that there is no evidence of improvement in the quality of services by the health facilities.[23] CHWs, who are socially motivated (as they traditionally are), were trained, and they benefited from some financial incentives of the program. Their training included, among other things, how to identify, monitor, and counsel pregnant women in their village areas, sell the vouchers, and to create demand for the vouchers by disseminating information regarding the vouchers. One possible explanation for why their impact may not have been strong is that the training may not have been good enough to enhance the quality of the service.

Additionally, in Table 2.6 Panel A (columns 9 and 10) — the 1st segment of the objective measure of quality — I note that there is no evidence for improvements, nor are there any differences between the program and non-program facilities in (a) the number of

[21] The possible explanation for the very good results on service provision by medical staff may be attributed to the fact that, from the onset, VSPs were assessed on their conformity to meet the clinical requirements for safe maternity services (their clinical governance and registration with government; technical competencies of the health workers; facilities, equipment and supplies management among others). The VMA indicated that only 2 awarded health facilities turned down the program (one on grounds of insufficient staff to meet the anticipated demand and the other failed to agree on the contracting terms (prices) for the services).

[22] Data from the VMA indicates that, for the term of the program, 15 VSPs were terminated due to consistent poor quality of services while 13 others were terminated due to misconduct like engaging in fraudulent activities

[23] Robust checks using sampling weights indicate evidence of improvement in services delivered by the health facilities as shown in Table 2A13 Panel B.

times that an inpatient was served; and (b) the inpatient waiting time before they received services from the medical personnel both during program implementation and after the program was withdrawn. This suggests that the increased demand for maternal health services in the program facilities did not worsen the quality of the services.

In the 2nd segment of the objective measure of quality, I considered improvements in the bed and staff numbers of the health facilities. The results shown in Table 2.7 indicate that during the program implementation period, the program increased the number of medical doctors, nurses, and midwives (indicated by the coefficients of program x post1, columns 10-12). Having said that, there were no differences in the number of beds between the program and non-program facilities (indicated by the coefficients of program x post1, columns 7-9). In the period following the program's withdrawal (post2), though there were no major differences in the dimensions for the objective measure of quality between the program and non-program facilities, there is evidence of an increase in the number of delivery beds and midwives at the program facilities, by 31.3% and 79.3%, respectively (indicated in the coefficients of program x post2). This positive effect on the number of delivery beds and midwives in Post2 is possibly to meet the demand for maternal care services seen in Table 2.5. Further, since there are no significant differences in the coefficients of program x post1 and program x post2 (row program x post1=program x post2, columns 7-12) for all these measures of quality, this suggests sustainable quality at the health facilities (measured by the number of beds and strength of medical staff), even after the program withdrawal. I supplemented this measure of quality by exploring the ratio of the number of medical staff (medical doctors, nurses, and midwives) to the number of beds (general beds, maternity beds

and delivery beds). Although the main results in Table 2.7 show a significant increase in the number of medical staff at the program facility during the program period, the results in Table 2A9 do not indicate such an improvement in the ratio of medical staff to beds among the program facilities although there is improvement in nurses to maternity beds and midwives to delivery beds in the program period as compared to the pre-program period (in row post1, columns 5 and 6). Taken together, there is no evidence of a decline in service quality at the program facilities, both during and after the program.

Collectively, the pattern suggests positive effects of the program on service quality at the program facilities, especially during the program implementation period. This effect does not endure after the program's withdrawal, except for the number of midwives and delivery beds. Since the increases in utilization of maternal care services and the service quality do not necessarily improve maternal and child health, the next subsection examines the program's impact on maternal and child health outcomes.

2.6.1.3 Impact on maternal and child health outcomes

Table 2.8 shows the estimation results for the maternal and child health outcomes. In Table 2.8 (Panel A), column 6, there is no evidence that the program decreased the incidences of complications during delivery (indicated in the coefficient of program_t). This being a self-reported measure, one wonders if women objectively know that they had a complication during the delivery in the absence of any indications by the medical personnel. Hence, I cannot say if the complications reported were serious or mild. To address this concern, or to ascertain if the program reduced the number of serious cases, I redefined the incidence of

birth complications as any complications including assisted and cesarean section deliveries. Based on this change, the results in column 7 still indicate no evidence that the program decreased the incidences of complications at birth. This may partly be due to the fact that the pregnancy data are from women who used the health facilities.

The results of child health outcomes also indicate no evidence that the URHVP had a significant positive effect on the birthweight or decreased the cases of low birthweight. Since 14% ((599-514)/599) of the children were not weighed at birth, and it is likely they were delivered at the health facility, I may not have captured children with low birthweights in this sample. To address this attrition bias, I did the same analyses by replacing the missing birthweights with the lowest and highest birthweights at each health facility, separately in the pre- and post-program periods. The estimation results shown in Table 2A10 are almost the same as the main results, which suggests that the attrition bias may not have been serious. Further, though the women survey is not nationally representative, the estimation model still controls for mother unobserved heterogeneity in addition to year fixed effects and district specific time trend. Even after all kinds of possible confounders were considered, there is no evidence that the child health, measured by infant birth weight, improved because of the program at the program facilities (indicated in row program_t, column 9 of Table 2.8).

Lastly, I supplement my findings on maternal and child health outcomes by using the pregnancy outcomes indicated in Table 2.8 (Panel B). Once again, I find no evidence that the URHVP had a significant positive effect on the likelihood of a pregnancy resulting in a livebirth (indicated in row program_t, column 5), nor a significant effect on the reduction of

incidences of stillbirths, miscarriages and neonatal deaths (indicated in row program_t, columns 6-8). In sum, there is no evidence for the positive effects of the program on the maternal and child health outcomes.

2.6.2 HETEROGENOUS EFFECTS

The impact of the program on the usage and quality of maternal health services may be larger in the districts with poorer households. This is because they are initially worse-off in health outcomes than the rest. Therefore, I have estimated heterogenous treatment effects based on the initial economic conditions (measured by the district's per capita GDP)[24]. I divide the subnational level data into two groups — districts that are above and below the median district-level per capita GDP of USD 196.50. I define the poorer districts with low per capita GDP as those whose per capita GDP is lower than the median value, and construct a dummy variable for the poor districts ($Poor_d$). I then estimate the models below for health-facility level (equation 2.4a for MoH data and 2.4b for Health facility survey data) and pregnancy level (equation 2.5) analyses.

(2.4a) $$Y_{jdt} = \delta_0 + \delta_1 P^1_{jdt} + \delta_2 P^2_{jdt} + \delta_3 (P^1_{jdt} \times Poor_d) + \delta_4 (P^2_{jdt} \times Poor_d) + \gamma_{dt} + \phi_j + T_t + \varepsilon_{jdt}$$

[24] I used the subnational level GDP data estimated by the Enhanced Light Intensity Model (Rafa M. et al., 2017). This model uses night-time lights as a representation of higher value-add economic activities and agricultural production — important aspects of Uganda's economy to provide a broad-based estimate of economic activity at the subnational level. The sub national data obtained is based on the Uganda national GDP data for 2014, a year prior to commencement of the URHVP.

(2.4b) $Y_{jdt} = \delta_0 + \delta_1 Post1_t + \delta_2 Post2_t + \delta_3(Poor_d \times Post1_t) + \delta_4(Poor_d \times Post2_t) + \delta_5(P_j \times Post1_t) + \delta_6(P_j \times Post2_t) + \delta_7(P_j \times Post1_t \times Poor_d) + \delta_8(P_j \times Post2_t \times Poor_d) + \gamma_{dt} + \varphi_j + \varepsilon_{jdt}$

(2.5) $Y_{ijdt} = \beta_0 + \beta_1 P_{jdt} + \beta_2(P_{jdt} \times Poor_d) + W_{ijdt} + \gamma_{dt} + \mu_i + T_t + \varepsilon_{ijdt}$

In equations 2.4a and 2.4b, the coefficient δ_3 (or δ_4) and δ_7 (or δ_8) capture the differential treatment effect of the program in the poor districts compared to the better-off districts during the program implementation (or after program withdrawal) respectively. Similarly, in equation 2.5, β_2 is the differential impact of the program on women in poorer districts compared with that on women in better-off districts.

The results of the heterogenous impact analysis of the program by district GDP are presented in Tables 2.9-2.12. As the coefficient of the interaction term (Poor x Program x Post2) is significant for the number of medical doctors at the health facility in the program withdrawal period (Table 2.11, column 4), this suggests that the program has pro-poor effects on this outcome. There is no evidence that the effect of the program on most of the outcome variables (maternal health service usage and quality of services in health facilities, women's and newborn babies' health outcome) is larger in the poor districts than the better-off districts. Therefore, it is suggested that the program's positive effects on the usage and service quality of facility delivery at the health facilities were not concentrated in wealthier districts.

2.6.3 ROBUSTNESS CHECKS

Using the MoH administrative data to estimate the average treatment effect of the program on the usage of maternal care services, I considered the annual number of women receiving services during the program (post 1) and after the program (post 2) taking into account the indicator for facilities that were under the program in year t, the year fixed effects, facility fixed effects, and the district specific time trends. Using the same approach to conduct a robustness check, I exploited the varying program intensities across health facilities for the average treatment effect on the usage of maternal care services by considering the panel facility fixed effects model.

$$Y_{jdt} = \delta_0 + \delta_1 P_t + \gamma_{dt} + \varphi_j + T_t + \varepsilon_{jdt}, \quad (2.6)$$

where P_t is the indicator variable that equals 1 for facilities that were under the program in year t; γ_{dt} represents the district specific time trends; φ_j is the facility fixed effects; T_t is the year fixed effects; and δ_1 is the coefficient of interest for the average treatment effect of the program. The results are shown in Table 2A11. Although these results do not provide for the analysis of post2 period, there is evidence that the program increased women's utilization of maternal care services (antenatal (4th visit), delivery, and postnatal services), which is consistent with the main findings in Table 2.5 (Panel A).

Secondly, since the health facility and women surveys are not nationally representative, I conducted further robustness checks of all the analyses by using sampling weights to

account for over-sampling of the program facilities[25]. The estimation results shown in Tables 2A12-2A15 present qualitatively similar findings as those obtained earlier in Table 2.5-2.8, except for those in Table 2A12 and Table 2A13 (Panel B). These findings indicate the significant usage of all maternal care services after the program's withdrawal and improvement in quality of services at the health facilities respectively. Therefore, the main results are less likely to be driven by the oversampling of the program facilities.

2.7 CONCLUSION

This chapter investigated how the URHVP support affects the quantity of maternal care use, quality of maternal care services, and health outcomes for mothers and children during and after the program's implementation. I utilized the annual administrative health facility MoH HMIS data and the survey data that I collected from both, the health facilities and the women who used the health facilities, to examine the impacts of the program on i) the quantity of maternal care use, ii) the quality of maternal care services, and iii) the health outcomes for mothers and children. By applying the difference-in-differences approach with health facility fixed effects and mother fixed effects, I show that the maternal voucher program enhanced the usage of maternal care services (ANC4, delivery, and PNC), the quality of these services, and the number of doctors, nurses, and midwives during the

[25] I compute the district-level weights for the non-program and program facilities as follows:

District-level weight for district s for non-program facilities = $\frac{total\ number\ of\ non-program\ facilities\ in\ district\ s}{total\ number\ of\ sampled\ non-program\ facilities\ in\ district\ s}$

District-level weight for district s for program facilities = $\frac{total\ number\ of\ \ program\ facilities\ in\ district\ s}{total\ number\ of\ sampled\ program\ facilities\ in\ district\ s}$

program's implementation period. However, these effects did not last after the program was withdrawn, except for the number of midwives. Further, even though the usage of maternal care services was enhanced during the program, there is no evidence that the URHVP improved child health outcomes — such as the infant's birthweight — or reduced complications during delivery, or improved pregnancy outcomes — such as an increase in livebirths, or a reduction in stillbirths, miscarriages and neonatal deaths. I also extended my analysis to investigate if the treatment effects were greater in poorer districts than in the better-off districts, as measured by the district's per capita GDP. Such pro-poor program effects were only found for the number of medical doctors in the health facilities after the program was withdrawn.

The URHVP not only provided maternal vouchers to mothers for free access to maternal care services, but also aimed to build the national health capacity through training, mentorship, and supportive supervision of the participating healthcare providers in partnership with the health teams. As reviewed in Hunter BM. et al. (2017), many studies have reported that maternal vouchers increased the use of antenatal care and a skilled attendant, births in healthcare facilities, and postnatal care. However, these programs have largely emphasized demand-side financing to increase the utilization of maternal care services with no supply-side components (except for payments for the care provided) to meet the increased demand. Hence, conclusions cannot be drawn about their effects on the quality of care. This study demonstrates that the URHVP increased the number of medical staff at the program facilities and improved the service quality provided by the health facilities and medical staff.

As a result of free maternal care services, the likelihood of women to seek skilled maternal healthcare may increase, though, they may still not receive the necessary or proper medical care from the providers owing to various constraints that vary between different service providers. According to the interviews of the health care providers on general concerns after the program, as indicated in Table 2A7, the main constraint in service provision was limited funding. Non-program health facilities considered the lack of equipment and work overload of their staff as the main reasons for their service limitations, while these were relatively minor reasons for the program facilities. Meanwhile, the service users (women) at the non-program facilities considered frequent stockouts of medical supplies (medicines and medical sundries), and the lack of equipment as the main reasons for limited service provision. On the contrary, service users in the program facilities did not consider these factors as problems. Since these questions on the issues in service provision were about the current situation at the time of survey, this may suggest that the program relaxed the constraints on medical supplies, equipment, and medical staff at the program facilities.

There are, however, issues that need to be solved for program sustainability. The results presented here suggest no evidence that the impact of the program on the number of doctors and nurses, as well as the quality of antenatal, facility delivery, and postnatal services remained after the program's withdrawal. Therefore, sustaining the positive effect of the program is an urgent and crucial problem for the design and implementation of future programs. Furthermore, I did not find evidence that the program improved maternal and child health outcomes. Finally, according to the heterogeneity analyses, the program could be

considered pro-poor only in one outcome. Thus, there is no evidence that the effects of the program on most of the outcomes were larger in the better-off districts than in poorer districts.

There are multiple limitations in this study. First, the surveys sought retrospective data, before the program (period 2011-2014), during the program (period 2015-2019) and after program withdrawal (6 months period 2019-2020). Although these periods were distinctive and respondents could easily remember the data requested, I acknowledge the possibility of recall error in the responses to retrospective questions. Second, as the selection of the women was based on those who had recently delivered multiple children, and given that I relied on the active CHWs for sampling these women, the findings may not be generalizable to all mothers due to selection limitations. Therefore, I suggest future analysis of the effect on maternal care utilization and child and maternal outcomes using randomly selected mothers. Third, I could not conduct individual-level heterogeneity analyses to examine if poorer women benefited more than the better-off women because I did not collect income data from the pregnant women survey. I estimated this heterogeneous treatment effect by using the district-level per capita GDP to identify poorer and better-off households. However, since the program primarily targeted poor pregnant women who were incapable of meeting their out-of-pocket expenses for maternal care, the heterogenous effects of the program, which are investigated by using the household-level income, should provide better policy implications. Fourth, the MoH data contained only one year (2014) in the pre-program period. I therefore estimated the pre-program trend using data from 2014 and 2015, on the premise that the program implementation was not fully active in 2015. This was because only 25.49% of the health facilities had received the program during the starting year. This data

limitation did not allow me to test the parallel trends assumption for the health facilities using multiple pre-intervention periods. I have requested the MoH for pre-program data covering at least 5 years and will run the test as soon as I receive this data. Fifth, as participating facilities started to receive the program at different points, 25.49% in 2015, 68.63% in 2016 and 5.88% in 2017, there may be concerns for the influence of some exogenous factors in this rollout. A comparison of facilities served in 2015 (treated group) versus those that started from 2016 (control group) and those served in 2015 and 2016 (treated group) versus those that started in 2017 (control group), detected no effect probably because of the small number (only 51 in total) of program facilities. There is need to examine this effect in future with a bigger number of program facilities. Also, I acknowledge that the average data collected for "during program period" under the Health Facility survey has some limitation ($Post1_t$ does not refer to the same calendar periods across facilities). In future, I should devise some ways to control for this. In addition, future revisions will also take into account the possible bias of using DID for treatment implementation at the multiple time periods as pointed out by Callaway and Sant'Anna (2021) and others. Sixth, I conducted the survey just 6 months after the closure of the program. Hence, the measurements for the program withdrawal period are only about 6 months long. Since the program facilities may have undergone some adjustments after the program was withdrawn, the estimated impacts in this study might be underestimated. To examine the long-term sustainability of the program, this period is too short to provide a true reflection of the program's long-term impact. Therefore, there is a need for the future analysis of this effect after a longer period. Finally, this study adopted the self-reported measures used in other studies. Since self-reported health measures may not

always be accurate, examining other indicators based on blood tests and medical records would be a fruitful future study.

CHAPTER 3

ESTIMATING THE EFFECT OF FINANCIAL INCENTIVES ON COMMUNITY HEALTH WORKERS' EFFORTS: EVIDENCE FROM A MATERNAL VOUCHER PROGRAM IN RURAL UGANDA

3.1 INTRODUCTION

In many developing countries, rural populations, particularly, still lack access to health services (Deserranno, 2019). To close this gap, many governments have widely utilized Community Health Workers (CHWs) as a popular community-based strategy to reach these populations and provide basic public services. CHWs have been integrated with the local health care systems and programs to improve healthcare service utilization and outcomes. These workers belong to the community of households where their services are provided. Thus, their activities are mainly conducted through household visits within the community, and include educating households on essential health behaviors, providing basic health care or advice, providing health-related products and referring severe cases to the formal health care system (Nyqvist et al., 2019). In the quest to achieve SDGs, the roles of CHWs have been expanded, and it is a matter of keen interest for the global community that the CHWs and their programs function effectively (WHO, 2016). However, many CHWs fall short of their assigned duties or activities due to multiple factors including low motivation and lack of inducements (Strachan et al., 2012).

As the motivation of CHWs is still a concern, understanding how to enthuse them is essential. Studies have proliferated in recent times focusing on what motivates community agents and social workers to contribute to the transfer of community services. However, the findings of these studies seem to point to disparate conclusions. Regarding the effect of performance pay on performance, some studies (Miller et al., 2012; Basinga et al., 2011) argue that financial incentives or entrepreneurial model programs increase the performance of agents and consequently, the usage of public services, thus enabling a substantial health impact.[26] Having said that, since the people that opt for community health work are expected to be socially motivated, it may be safe to assume that providing free service would be more socially rewarding than the financial incentives of the entrepreneurial model. In fact, some studies have demonstrated that pro-social CHWs function better than monetarily driven ones (Ashraf et al., 2014; Wagner, Z. et al., 2019). Nonetheless, as already seen, this may not imply that entrepreneurial models are ineffective. They may be more effective for health commodities that have a higher markup per sale or for households that are less price sensitive. They may also work more effectively if enhanced with regular training, monitoring, target reviews, provision of bonuses to reward performance, enforcement of penalties, and dismissal for failure to meet targets and misconduct, among others. As to how sustainable financial incentives are, Celhay et al. (2019) find that short-term inducements to health

[26] Financial incentives are referred to as any type of financial reward that motivates agents to work harder such as higher wages and higher performance reward among others (Deserranno 2019) while the entrepreneurial model is one where agents sell health commodities door-to-door and retain the profits" (Nyqvist et al., 2019; Wagner, Z. et al., 2019).

providers (clinics) increased innovations, which in turn raised the rate of early commencement of prenatal care. These improvements persisted while the incentives lasted and for at least 24 months after the incentives ended. Therefore, depending on the circumstances, social workers could respond differently to social incentives, financial incentives or both.

Although previous studies have attempted to investigate whether financially motivated or socially motivated agents are more effective in improving healthcare service utilization, what has remained unanswered is if CHW services under the entrepreneurial model crowd out other health services, which may negatively affect people with health problems other than the incentivized services, and if the CHWs' mentality for volunteer work remains the same once the incentive is no longer provided. The answer to both these questions may result in negative unintended consequences. Since CHWs are socially motivated people and are usually engaged in various activities within the community (volunteering, charity work and political participation), financial incentives to CHW services can crowd out such activities. It is therefore crucial to understand the likelihood of a crowding out effect as a result of the introduction of incentivized programs.

I take a case of the Uganda Reproductive Health Voucher Program (URHVP) to examine the possibility of these trade-offs. Previously, the CHWs in these program health facilities did not receive any financial rewards from the government for their services to the community. Now, under this program, they are entitled to financial incentives: i) a transport allowance being the markup from the sale of each voucher (used for transport purposes or as

additional income for those CHW who use a bicycle or walk); and ii) a monthly bonus payment if they followed-up with their voucher clients to ensure the voucher was used. These financial incentives were aimed at boosting CHW efforts in the delivery of program services by reducing the challenges of low motivation and lack of transportation to make their visits, as had been the case. Not all activities were incentivized by the program, such as the treatment of common and uncomplicated diseases (pneumonia, worm infestations, diarrhea etc.), and provision of HIV/AIDS preventive education along with counselling services.[27]

I conducted a survey of 272 CHWs at 140 health facilities across 30 districts, and collected panel datasets which cover the periods before, during, and after the program. I estimated the program's impact on CHW efforts using the Difference-in-Differences (DID) approach with fixed effects estimation. The empirical analyses show that the program increased CHW service delivery in terms of the number of villages and households covered, and increased working hours per week. However, I did not find evidence that non-incentivized services were crowded out due to the introduction of URHVP. This is despite the program CHWs decreasing their working hours per week for other economic activities to increase their service delivery. Unlike the study by Celhay et al. (2019), the effect of the URHVP on the efforts of the CHWs did not persist after the withdrawal of the incentives.

These results contribute to the broad literature that assesses the effectiveness of financial

[27] In accordance with MoH Village Health Team Strategy and Operational Guidelines, 2010, and the URHVP CHW guidelines, I consider the URHVP incentivized services by CHWs to include sensitization/health education, mobilization of communities to utilize health services during health campaigns, referral of cases from villages to health facilities, data collection and reporting during community disease surveillance and the counselling of women among others.

incentives or entrepreneurial models on the performance of CHWs and their motivation for public service delivery. I provide a basis for informed policy interventions in CHW programs that are aimed at improving healthcare service utilization, service provision, and health outcomes. Most of the related literature focuses on the influence of earnings levels on the traits of applicants for public service jobs (Da Bo et al., 2013; Ashraf et al., 2020; Deserrano 2019); the impact of performance pay on the performance of agents for health service delivery (Ashraf et al., 2014; Miller et al., 2012; Celhay et al. 2019; Wagner, Z. et al. 2019) or education (Fryer 2013; Duflo et al., 2012). In the context of this literature, my contribution is to provide evidence that an incentivized program not only impacts the effort made by CHWs, but also their work ethic and the provision of non-incentivized services, both during the incentive and after its withdrawal.

The rest of the chapter is organized as follows: Section 3.2 presents the literature review followed by the context on CHWs in Uganda, while Section 3.3 provides a description of the voucher program. Section 3.4 describes the data used in the analyses, while Section 3.5 presents the empirical models used. Section 3.6 presents the estimation results and discussion, while Section 3.7 concludes this chapter.

3.2 Literature Review

The effects of incentives on public service providers have largely been explored in experimental studies, which have shown that individuals respond to financial incentives in different ways. These studies have focused on incentives given for health, education, and in some cases, agriculture service provision. This section focuses on the effects incentives have

on the provision of health services. In this context, previous studies have examined the impact of earnings levels on the behavior of applicants for public service jobs (Ashraf et al., 2020; Deserranno, 2019); the effect of incentives on service provider attendance (Banerjee et al., 2008); the effect of price-cap regulations on the social welfare of agents (Blum 2020); the effect of incentives on health outcomes (Miller et al., 2012); and the effect of incentives on the performance of agents when delivering health services (Ashraf et al., 2014; Celhay et al. 2019; Wagner et al., 2019).

Financial incentives can change the type of applicants who enter the public sector — a sector where motivation toward public service is a critical characteristic to participate effectively. Ashraf et al. (2020) embedded experimental variations in a job advertisement in Zambia's nationwide recruitment drive to hire 330 community health care workers. In some districts, the job advertisement highlighted the prospect of career advancement (career incentives), while in other districts, the advertisement stressed the social importance of the job (social incentives). The authors, thus, set out to test if career incentives would attract talent at the expense of prosocial motivations. The results indicate that offering career opportunities for community-based jobs attracts more significantly qualified applicants who also have a high degree of prosocial motivation. Thus, career incentives do not crowd out prosocially motivated applicants for higher quality ones, implying that career motivated recruits are more talented, equally prosocial, and able to deliver health services with exceptional health impact.

On the contrary, Deserranno (2019) finds that monetary incentives can lead to a less socially motivated applicant pool. The author created experimental variations in the expected earnings to estimate the impact of monetary incentives on candidates' perceptions of a health promoter job in Uganda's rural villages, and on size and makeup of the applicant pool. In one treatment group, the job advertisement mentioned a minimum amount the health promoter was expected to earn (low-pay), while in the other treatment group, the job advertisement mentioned a maximum amount the health promoter could expect to earn (high-pay), and in the 3rd treatment group, the job advertisement mentioned the mean of the expected earnings distribution (medium-pay). Compared with the low-pay treatment group, the high pay treatment group attracted 30% more applicants. However, they also had less experience as volunteers and less prosociality. Specifically, experimental units in the high pay group donated 55% less than those in low-pay group. They were also 32% less likely to volunteer in the health sector. Results further indicate that money can disincentivize pro-social behavior at the application stage since it discourages such volunteers from applying to the job (and yet they are found to be more committed to the job and are better performers). Further, Deserranno (2019) shows that prosocial behavior and job performance are correlated. When compared to the units in the high pay group, those in the low-pay group were found to be more prosocial and had a higher aggregate performance in their 1st year of work. They also visited more households, arranged for more demonstrations in the village, and were more likely to target the most vulnerable households.

Health workers in government clinics tend to have no threats of being fired, even without attending official hours. Banerjee et al. (2008) examined the impact of a financial

penalty (reduced salary) for those who are not present during their office hours. They hoped to investigate if such penalties are effective to solve the problem of absenteeism. In this randomized experiment in India, the government established a monitoring system aimed at increasing Assistant Nurse Midwife (ANM) attendance levels, and in turn, to punish nurses for their absenteeism in the public health facilities (rural subcenters). Protected time/date stamp machines (timeclocks) were used to verify and monitor ANM attendance. Nurses who were noted as absent for more than 50% of the days in a month had their wages reduced by the number of days they were absent, while those who were absent for more than 50% of the days per month for two consecutive months were suspended from service. Initially (6 months into the program), the system was effective as there was a substantial treatment effect. However, after a few months, the scheme was undermined by the local health administration, as nurses were able to claim an increasing number of exempt days from work. As such, the effect diminished over time and was zero (ineffective) at the end of the study. Therefore, 16 months after program inception, there was no difference in the rates of absence between the treatment and comparison centers. Clearly, like other public service agents, nurses are receptive to properly administered incentives. The difference here is that in Duflo et al. (2012)[28], the incentives may have been enforced by the NGO that was running the schools, which was not the case in Banerjee et al. (2008). Here, the government did not have sufficient independence to enforce the incentives.

[28] Duflo et al. (2012) found that after 1 year, teachers' attendance improved in schools where teachers were incentivized, as did their students' learning. This led to their test scores being 0.17 standard deviations higher than those in the control schools.

Thus, what type of incentives are effective at enhancing service delivery? Ashraf et al. (2014) conducted experiments in Zambia by giving financial and non-financial incentives to public health workers (agents) to sell female condoms in a move to promote HIV prevention. They were randomly assigned monetary rewards (with margins of 10% or 90% on each condom sale) or non-monetary rewards, where agents received stars for each sale. Agents in the control group, however, were enrolled as volunteers and received no incentives, monetary or otherwise. Evidence from this study indicates that i) agents in the non-financial rewards arm were more effective at selling condoms, selling twice as many as those in the financial rewards treatment group; ii) both types of rewards had a stronger impact on the prosocially motivated agents; and iii) both types of rewards were effective if their relative value was high. The authors indicate that financial incentives may have been ineffective because the demand for condoms and earnings from each sale were too low.

In another similar study, Wagner et al. (2019) conducted a field experiment in Uganda to examine if an entrepreneurial model increased the effort made by CHWs (measured by home visits made) relative to a free health product distribution. CHWs were randomly assigned to either sell Oral Rehydration Salts and Zinc (ORS & Zinc) — a treatment for child diarrhea — door-to-door and retain the profits, or to distribute the treatment at homes for free. Compared to the free distribution, the entrepreneurial model led to substantially less CHW effort, that is, 35% households visited versus 61%. Qualitative interviews indicate that since the CHWs served within the communities they lived in and were known, selling of health products may have attracted a social penalty, while free distribution was socially rewarding. Although I too study the impact on CHW efforts toward

service delivery, the two studies look at completely different programs. Wagner et al. introduce a social cost (resentment in the community) that later weakens the impact of CHW efforts in the community. The URHVP (full cost maternal care vs. subsidized price), on the other hand, provides an opportunity for poor mothers to buy a voucher for free services, and therefore, expect a better service than the status quo (having to pay for all services).

As indicated by Finan et al. (2017), financial incentives do matter for agents as they may increase performance along the incentivized dimension. However, it is important to note that the nature of the incentives matters most for performance. Incentives often come with tradeoffs, such as the discouragement of effort among prosocially motivated people, and multitasking between social/public work and economic activities. There is a dearth of empirical studies that examine these tradeoffs induced by financial incentives. This study attempts to address this concern by examining if incentives crowd out CHW effort toward non-incentivized services and other community activities, and if these incentives affect the work ethic of CHWs after the program's withdrawal.

3.3 CONTEXT

3.3.1 CHWS IN UGANDA

Uganda's Ministry of Health (MoH) developed the CHW (popularly known as the Village Health Team (VHT)) strategy in 2001 to empower community participation. It aimed to improve community health and strengthen the delivery of health services at the community

and household levels[29]. This approach was in line with the Alma Ata (1978)[30] and Ouagadougou (2008)[31] Declarations on Primary Health Care. According to the operational guidelines of the CHW strategy, the CHW is a non-statutory community (village) structure, where the workers are selected by the community members. They are expected to manage matters related to health and cross-cutting issues aimed at promoting the health and well-being of all the village members. Among the criteria for selection, a CHW should be willing to serve as a volunteer, be a resident of the village, be available to perform specified CHW tasks, be interested in health and development matters, be a good mobilizer and communicator, be able to read and write in the local dialect at least, be dependable, approachable, a good listener, and be 18 years or older. Some CHWs may be selected by the district/community leaders or health officials or program implementing partners. The CHW strategy relies heavily on the concept of volunteerism.

The CHWs' tasks are largely in the core areas of primary health care[32]. These include: community information management, health promotion and education (sensitization on good hygiene and sanitation practices, such as washing hands, using pit latrines and drinking boiled water; importance of seeking health services, like HIV testing and counselling), organizing

[29] National Village Health Teams assessment in Uganda, Ministry of Health, Uganda – 2015; Community Health Extension Workers Policy, 2018.
[30] Alma Ata (1978) Declaration on Primary Health Care. Retrieved on November 11, 2021 https://www.who.int/teams/social-determinants-of-health/declaration-of-alma-ata.
[31] Ouagadougou (2008) Declaration on Primary Health Care. Retrieved on November 11, 2021, https://www.afro.who.int/publications/ouagadougou-declaration-primary-health-care-and-health-systems-africa.
[32] MoH Village Health Team Strategy and Operational Guidelines, 2010. Retrieved on August 10, 2021, http://library.health.go.ug/publications/health-education/village-health-team-strategy-and-operational-guidelines.

communities to use preventive health services (immunization, use of mosquito nets etc.), management and treatment of common and uncomplicated illnesses (malaria, diarrhea, pneumonia, worm infestations, etc.), data collection and reporting during disease surveillance, distribution of health commodities (such as deworming and anti-malaria tablets), women counselling, referral of cases to health facilities, following-up with discharged patients and those on long-term treatment.

The implementation, coordination, monitoring and supervision of the CHW strategy is directed at the national level, while the activities of CHWs are supported at the district level. Based on the MoH guidelines, CHWs are initially trained by the district CHW Trainers using the Training of Trainers (TOT) approach. This training involves preventing diseases, educating on health, profiling the community, creating records for the community, home visits, organizing the community, and making referrals. Refresher trainings are occasionally provided, and these vary in their content, duration and methodology. The supervision of CHW activities is done by the District Health Teams, District leaders and their Implementing Partners (IPs). However, the reporting is non-uniform or lacking in many districts.

The government funds the CHW strategy through the MoH and several IPs. The IPs work with the districts to support CHW activities. The IPs use the MoH guidelines for the implementation of CHW activities to provide financial, technical, and logistical support to CHW programs. While they largely provide monetary motivation, they also provide non-financial forms of motivation, such as verbal recognition at public gatherings and in the media; capacity building; and supplies such as uniforms, bags, gumboots, umbrellas, identity

cards, and bicycles. However, these IPs are not equally distributed geographically; their activities are concentrated in some districts and not in others.

It is difficult to monitor CHWs because they have large level of autonomy and flexible working hours, while often working part-time in addition to their other self-managed activities, such as farming. Although the CHW strategy in Uganda is still characterized by its challenges, such as the lack of financing, weak coordination structure at all levels, poor monitoring and supervision, and inefficient training, the functions of the CHWs are very critical to close the gaps in the delivery of health services to the community (Deserranno, 2019).

3.3.2 CHWs UNDER THE URHVP

The URHVP recruited a group of active CHWs (as recommended) from a pool of existing CHWs at the District health office to aid with the program's implementation. For the voucher program, the CHWs were recruited from the district pool by the Voucher Management Agency (VMA) and District Health Team (DHT), attaching at least 2 CHWs to each program health facility. The criteria for recruitment into the voucher program were as follows: previous training in maternal health; good conduct and a credibility report from the Local Council 1 (LCI)[33]; the ability to read, speak, and write basic English and a local dialect; and lastly, the interest to participate in the voucher program. The CHWs were trained on the voucher program, how to use the poverty grading tool, and how to report voucher sales using

[33] LCI is the lowest local government office of a village level, which is the lowest political administrative unit in Uganda. This is preceded by the parish run by an LCII, subcounty run by an LCIII, County run by an LCIV and district run by an LCV (UBOS website).

the MarieTXT SMS system. The primary responsibilities of the CHWs were to sell vouchers and to create demand for them by disseminating information. They had to identify, monitor, and counsel pregnant women in their respective village areas.

The selection of the primary beneficiaries was done by the CHWs. Pregnant women were identified or located through the village structures and social gatherings. The CHW attended all such gatherings and events to disseminate information regarding the voucher program. Once pregnant women were known by anybody in the community, the CHW was alerted. The CHW then visited the women's homes and used a customized poverty grading tool (face-to-face survey) to collect data on demographics, current pregnancy status, prior births, household welfare and assets, and other factors. This information was used to assess the poverty level of the pregnant woman. Only those women who scored 12 points out of 21, or lower, were considered poor and eligible to purchase the voucher. Once identified as eligible, the women were informed of the potential benefits of the program and were invited to purchase the voucher, upon consent from their family.

The vouchers were sold to the CHWs at UGX 2,700 (USD 0.73)[34] per voucher and the CHWs in turn sold the vouchers to the eligible women at UGX 4,000 (USD 1.09). The markup of UGX 1,300 (USD 0.35) was used as a transport allowance for the CHW, or an income for those CHWs who could use a bicycle or walk. The eligible pregnant women who purchased the voucher were given a hardcopy of the voucher, which had a unique ID number

[34] Using exchange rate for September 01, 2015: USD1 = UGX3,676.36, retrieved from https://www.bou.or.ug/bou/collateral/interbank_forms/2015/Sep/Major_01Sep2015.html

on it. They were also given information about all the participating Voucher Service Providers (VSPs) within their catchment area, and told to present the voucher at the health facilities when seeking any services under the voucher package. The VMA used the MarieTXT SMS system to monitor the voucher distribution and sales by the CHW, and service provision by the VSP. The CHW were required to follow up with their voucher clients to ensure usage of the voucher; redemption of over 80% per month was rewarded with UGX 30,000 (USD 18.16) per month to the CHW.[35] This is a significant boost to the income of people in the rural areas of Uganda. The CHWs were regularly assessed to ensure the provision of quality care and adherence to the service guidelines issued by the VMA and the DHT.

Taken together, the primary roles of the CHWs under the URHVP included sensitization of pregnant women, selling of vouchers, follow-up of women to use voucher services (antenatal care, delivery, and postnatal care), follow-up of women to use non-voucher services (immunization of newborn babies), and submission of reports on the program. Accordingly, I consider sensitization/health education, mobilization during health campaigns, referral of cases from villages to health facilities, data collection and reporting,

[35] When a CHW sold a voucher, he/she entered the data in the MarieTXT system. Also, when a woman visited the health facility, the visit was also captured in the MarieTXT system. This system was used for monitoring voucher distribution, keep track of voucher sales by CHWs and service provision at the health facilities and also for computing the redemption rate. The redemption rate of >80% was calculated from the total number of vouchers for each CHW across all services ANC, delivery and PNC. Women pregnancies were at different stages so mothers sought different services at different times (when they were scheduled or supposed to do so except for emergencies). It was therefore not likely that CHWs would push mothers to go to health facilities when they are not supposed to do so. The motivation for the redemption strategy was to encourage CHWs to market the voucher and sell as many vouchers as possible but most importantly to follow-up on the usage of voucher services.

women counselling, and other services (immunization of babies) as the incentivized health services under the URHVP while the treatment of uncomplicated diseases (pneumonia, worm infestations, diarrhea etc.), provision of HIV/AIDS preventive education and counselling services, as non-incentivized by the URHVP.

3.3.3 RESEARCH QUESTIONS

To deliver its core objective of increasing women's utilization of skilled maternal care services among the rural poor women, the URHVP employed the services of CHWs to identify and sell vouchers to program beneficiaries. As CHW services are constrained due to low motivation and lack of inducements among others (Strachan et al., 2012), the URHVP provided financial incentives to them, to motivate their efforts. Thus, this study's objective is to examine the effects of temporary financial incentives on the CHWs efforts for service provision during and after the program, and identifying any crowding out effects on the non-incentivized services.

First, I predict that URHVP financial incentives improved CHW efforts during and after the program, as measured by their coverage (the number of villages and households) and time spent on CHW related work. My concern is that the program might decrease the CHW's efforts once the financial incentive is taken away. This is analyzed by measuring the impact on the CHWs' efforts after the program ends.

Second, as the program financially incentivizes them to work more on URHVP services, I examine the prediction that, CHWs decrease their time in the provision of services that are not considered for URHVP payment. If that is the case, it might be concluded that the program crowds out other services to maximize focus on the program related services.

This would, understandably, negatively affect people with health problems other than maternal health. Further, the people who work as CHWs are engaged in other cross-cutting activities to support their communities. Therefore, I investigate my prediction, that the program financially incentivizes them to work more on incentivized services, which in turn decreases their involvement in other community work (prosocial behavior activities- charity, volunteer and political party). Such a change, may not be a good thing for their communities. Thus, I examine if such a crowding out effect triggered by the introduction of URHVP is found.

I, therefore, expect to answer these research questions: a) did the URHVP improve CHWs' service delivery in terms of their coverage and time spent on CHW work during and after program? b) did the URHVP crowd out CHWs' efforts to provide non-incentivized CHW services and remain involved in community activities (charity, volunteer and political party) by focusing exclusively on program related services?

The Uganda Demographic Health Survey 2016 indicates that people travel an average of about 5km to access health care. This is particularly true in the rural areas of Uganda. To bridge this immense gap, CHWs trek long distances to ensure access to, and the continuity of health service delivery to their communities amidst various challenges as shown in Table 3A3. Therefore, their critical role as the first line of contact and essential workers cannot be over-emphasized. I note that 73% and 64% of CHWs indicated the lack of transport options, and having to travel long distances, respectively, as their greatest challenge in the execution

of their social work.[36] CHWs who previously did not receive any financial rewards from government for their services to the community, received financial incentives in terms of transport allowances and bonus, under the URHVP. It is in this context that I expect the incentive provided to the program CHWs, in terms of transport allowance, to increase their effort with regard to the number of villages and households visited during the program's implementation period. However, once this transportation allowance is withdrawn after the program's completion, CHWs may once again be unable to move around to provide services.

3.4 Data and Descriptive Statistics

3.4.1 URHVP CHW Survey

Data for my main analyses comes from the URHVP CHW survey that I conducted between January and March 2020. Table 3A1 (Panel A and Panel B) provides detailed information on the sample for this study. The URHVP CHW survey was conducted at 140 health facilities across 30 districts, 3 regions (Central, East, and West); 12 districts did not receive the program while 18 districts did. The survey data constituted 51 program health facilities (supported by URHVP) and 89 non-program facilities (not supported by URHVP). The non-program facilities were selected from both program and non-program districts (23 and 66, respectively).

To sample the health facilities, I first selected 18 out of the 25 URHVP project districts (9 in the western region and 9 in the eastern region) based on the district health

[36] Consistently, in the sub sample, 75.4% and 66.5% indicate the lack of transport and travelling of long distances respectively as their greatest challenges.

indicator (if it was less than or equal to the national average of 44%) in 2014[37]. For non-program districts, I selected 12 districts (3 in the Western region, 3 in the Eastern region, and 6 in the Central region) that neighbored the program districts and had comparable maternal health conditions to those of the program districts. This was done by setting a threshold index level of 21% and above (i.e., the percentage of children delivered at maternity institutions out of all the children born in a given time). Second, from the URHVP project documents, I found that there were 110 program health facilities in the selected 18 districts. The program facilities were included in the program if they obtained a score of 65% and more, as an indication of their capacity to provide safe maternity services. There were no known exogenous factors that affected the selection process of program facilities into the program. Given a limited budget that allowed data collection from 140 health facilities, I purposely oversampled the program's health facilities to ensure statistical relevance. After sorting the program health facilities by district, subcounty, ownership, and levels (HCI, II, III, IV or clinics), I chose every second program health facility as the sample program health facility. Third, after gathering the list of all health facilities in my sampled program and the non-program districts from the Ministry of Health's Statistics Office (where there are 1,154 non-program health facilities) I sorted them by district, subcounty, ownership, and levels and sampled one out of every 13 non-program facilities. This resulted in the selection of 66 non-program facilities the from non-program districts, and 23 similar facilities from the program

[37] Dropped were outliers (Luuka District at 18% in the eastern region) and those that exceeded the national average (Bushenyi District (57%), Isingiro District (48%), Mbarara District (65%) from the western region and Busia District (54%), Iganga District (50%), Jinja District (65%) from the eastern region.

districts. The MoH and DHO granted me permission to conduct interviews with the Facility Heads (In-Charge), Mothers, and CHWs at their respective health facilities.

The sampling for the CHW's survey was done using the snowball technique. In each facility, I asked the facility in-charge to list some active CHWs who were attached to the facility for more than 5 years.[38] If the CHW had not supported the facility for at least 5 years, the previous CHW was selected for the interview of the earlier period. CHW respondents from the program facilities had to be URHVP participants and I selected 2 CHWs at each health facility. Data were collected using a structured questionnaire by CAPI (Survey Solutions). We conducted the URHVP CHW survey in English.[39] During the enumerator trainings, we conducted pilot and pre-test exercises to confirm the accuracy of the tools and the respondent's confidence in the recall data before the actual data collection commenced. The survey sought retrospective data from before the program (period 2011-2014), during the program (period 2015-2019), and after program withdrawal (6-month period in 2019-2020). As these periods were distinctive, respondents could easily remember the data requested. I however, acknowledge the possibility of recall error in the responses to

[38] At the program facilities, the VSP selected 2 or 4 most active CHWs per facility from an active pool of 8 or 10. In turn, for purposes of the survey, the health facility in-charge selected the most active 2 CHWs to participate in the survey. Similarly, at the non-program facilities, the health facility in-charge selected the most active 2 CHWs who would have participated in the program should the opportunity have been presented. As such, the CHWs selected for the survey were comparable and any differences seen are not from the selection process but otherwise

[39] In some other analyses of this paper, I use data from the URHVP Health Facility survey which was conducted in English and URHVP Women survey which was conducted in three local languages; Luganda for the Central region, Lusoga for the Eastern region and Runyakitara for the Western region.

retrospective questions. A team of 11 experienced and trained enumerators collected the data under my close supervision.

We interviewed 272 respondents (97 from the program facilities and 175 from the non-program facilities) on their age, marital status, religion, education, residence details, biological children, and dependents.[40] We also interviewed them about their CHW membership, experiences, training in the last 10 years, challenges, issues in service provision, public service motivation[41], leadership record, prosocial behavior[42], and personality. Data were collected for the three defined periods (pre-program, program implementation, and after program withdrawal). The data focused on factors such as occupation, assets, access to financial services, services provided, villages and households coverage, time use, service delivery at the facility, and prosocial behavior. Respondents from the program facilities were also interviewed on the URHVP program (selection for the program, expectations from the program, voucher sales, and perceptions about the program outcomes). I constructed a 3-period panel dataset from this information.

[40] The CHW Survey targeted 2 CHWs from each health facility (2×140=280). However, we had 5 fewer from the program facilities (Kibuku HCIV, Iyolwa HCIII, Kabwohe HCIV, Allied Health Medical Center Clinic and Bugongi HCII) and 3 fewer from the non-program facilities (Family Health Resource Center Clinic, Suubi Medical Center HCIII and Bakhita HCII).

[41] Six dimensions of PSM (Attraction to policy making; Commitment to public interest; Social justice; Civic duty; Compassion; and Self-sacrifice).

[42] Charity work (blood donation, fundraising, movement for social benefits and other charity work); volunteer work (agriculture extension work, awareness & advocacy programs, tree planting and other volunteer activities); and political party work (party member campaigns, party agent & vote counting, party mobilization campaigns, party communicator and other party activities).

3.4.2 DESCRIPTIVE STATISTICS

The tables below show the detailed summary statistics for the CHWs. I categorized them by program and non-program facilities.

Table 3.1 shows that there are no major differences between the CHWs in the program and non-program facilities, as most of the important aspects are comparable across both study groups. The first segment contains basic CHW information. On average, 53.3% are female; 47 years old; 91.5% are Christians; 97.4% have children and dependents; 6 biological children. 54.8% of the CHWs have lived in the same residence since birth, whereas those who were not born in their current residences have lived there for over 20 years. A higher share of CHWs are married or living with a partner in the program facilities (90.7%) than in the non-program facilities (78.3%). In the 2nd segment on CHW selection, duration, training, and participation in maternal health programs, I note that 73.5% of the CHWs are selected by the community popular vote; they have been CHWs for more than 12 years since their selection and have supported their current facilities for about 11 years; 95.2% received training on their appointment as a CHW, while 97.4% have received health related training in the last 10 years. In the 3rd segment, on education and information access, I note that the literacy level among CHWs is generally high. All CHWs have attended school with 90.1% completing at least primary level (this share being higher for the program facilities at 96.9%). This difference is not a cause for worry since the nature of their work requires CHWs to be able to read, write, and speak their local dialect. This important aspect is comparable across the two study groups. Radio is the most accessible source of information with 89.3% of the CHWs listening to it almost every day, or at least once a week. Newspaper and internet

consumption are generally very low, especially among CHWs from the non-program facilities. Similarly, as shown in Table 3A2, the study groups are largely comparable by their occupation, asset possession, and access to financial services. This is true across all periods, except for the possession of bicycles, which is greater for program CHWs than the non-program workers.

Tables 3.2 and 3.3 show t-test results comparing the program and non-program CHWs on work coverage, time use, healthcare service provision, and community prosocial behavior activities (charity work, volunteering and political party participation)[43] in the three periods — pre-program period (before URHVP), program implementation period (post1) and program withdrawal period (post2). I also show the changes over time. Columns 6 and 7 show, the significant differences ($p<0.05$) in the outcome variables of each study group between periods pre and post1 (marked [a]). Columns 10 and 11 show the significant differences in the outcome variables of each study group between periods pre and post2 (marked [b]) and periods post1 and post2 (marked [c]).

In Table 3.2, it is important to note that there are no significant differences in the main outcomes between the program and non-program CHWs in the pre-program period. The significant differences between the two groups of CHWs are only noted in the program period (post1). Specifically, when compared with the non-program CHWs, the program CHWs

[43] Charity activities (blood donation, fundraising, movement for social benefits, and other charity work); Volunteer activities (agricultural extension work, awareness & advocacy programs, tree planting, and other volunteer activities) and Political party activities (party member campaigns, party agent & vote counting, party mobilization campaigns, party communicator, and other party activities)

covered more villages and households, spent more hours per week[44] on CHW work and equally lesser hours on their 1st major economic activity, increased their effort in the provision of all incentivized services, and also provided more time for the non-incentivized services (HIV/AIDS). During the program period, there is no evidence of increased effort in service delivery among the non-program CHWs (column 6). By contrast, there was increased effort in service delivery (in terms of the factors mentioned above) by program CHWs in the program period. However, this effect vanishes after the withdrawal of the program. During the same period, more time (in hours per week) is devoted to the CHWs' economic activities such as farming. As reported by Finan et al. (2017), and as noted here, outside of the CHW work, farming activities dominate the workers' 1st major activities at 82.48%.[45]

In Table 3.3, the results indicate that there are largely no differences between the program and non-program CHWs' involvement in their community prosocial activities, except for the higher involvement of non-program CHWs in political activities in the pre-program period. When compared with the pre-program period, there is increased involvement by both program and non-program CHWs in all prosocial activities during the program period. This increase in involvement in community activities may be attributed to the many political activities that occurred during this time (Uganda held its general elections in 2016). In their search for political popularity and votes, politicians tend to engage voters in various charity, volunteer, and political activities. Meanwhile, involvement in political

[44] Hours per week taken by multiplying the days per week by the hours per day.
[45] CHWs were asked about their 2nd major activity away from CHW work, and almost 50% indicated that they had none. I do not consider the 2nd major activity for the analyses on CHWs time use.

party activities by both the program and non-program CHWs greatly deteriorated after the program's withdrawal.

From the above statistics, the notable differences are that compared to the non-program CHWs, their program counterparts are more likely to be married/living with a partner and have completed at least primary level education. Since these two characteristics are significantly different between the two groups, I created a sub sample by maintaining only those CHWs that are married/living with a partner, and who have completed at least primary education. The descriptive statistics of this sub sample indicate that there are no differences between the CHWs in the program and non-program facilities, and that all the other statistics discussed above are almost the same as those for the full sample (sub sample statistics are not shown for brevity).

3.5 EMPIRICAL MODEL

The descriptive statistics show that, as a result of the program, there was increased CHW effort and service delivery in terms of the number of villages and households covered, working hours per week (consequently, the hours spent on economic activities declined), provision of all incentivized services and non-incentivized services (HIV/AIDS), and participation in community work at the program facilities during the program period. However, some of these effects disappear after the withdrawal of the program.

In this section, I conduct regressions to examine the impact of the URHVP incentives on the efforts of CHWs toward service delivery, and the effect on non-incentivized services and

involvement in community activities, both during the program period and after withdrawal of the program. Table 3A1 (Panel C) provides detailed information on the data description for this study. The data was collected in average terms for the respective periods specified by before/pre- URHVP (2011-2014), during URHVP (2015-2019) and after URHVP ended (6 months 2019-2020). First, I explore service delivery by i) estimating the average number of villages and households supported by the CHWs; ii) the average number of hours per week each CHW spends on their work and their 1st major economic activity. Second, I estimate the impact of the program on the average number of days a month a CHW spends on the provision of incentivized health services and, more importantly, non-incentivized services (which were not covered by the URHVP). Lastly, since CHWs provide other cross-cutting community services, I also estimate whether their involvement in these community activities (charity work, volunteering, and political party participation) decreased due to the introduction of the URHVP. These analyses allow me to test if financial incentives crowd out social motivation and estimate their effect on the work ethic of the CHWs during the program's implementation (post1) and after the program's withdrawal (post2). The identification strategy is as follows: having controlled for any unobserved health facility heterogeneity and year fixed effects to control for common shocks, as well as district-specific linear time trends to account for the possibility that the services evolved differently in the program and control facilities, the treatment status of the program and the error term in the service provision model should not be correlated.

Using the CHW survey data, I employed the difference-in-differences CHW fixed effects model to investigate the average effects of URHVP during program implementation (post1) and after implementation (post2) using the following specification:

$$(3.1)\ Y_{ijdt} = \delta_0 + \delta_1 Post1_t + \delta_2 Post2_t + \delta_3(P_j \times Post1_t) + \delta_4(P_j \times Post2_t) + V_{ijdt} + \gamma_{dt} + \mu_i + \varepsilon_{ijdt},$$

where Y_{ijdt} indicates the following outcomes of interest for CHW *i* at health facility *j* in district *d* in period *t*: a) the natural log of the average number of villages and households supported by the CHW in period *t*; b) the natural log of the average number of days per week, hours per day and hours per week the CHW spends on CHW work and their 1st major economic activity in period *t*; c) the natural log of the average number of days a month the CHW spends on the provision of incentivized and non-incentivized health services in period *t*; and d) an indicator variable taking 1 if CHW *i* was involved in prosocial behavior/community activities in period *t*. $Post1_t$ and $Post2_t$ are indicator variables that equal 1 for the program-implementation period and after program withdrawal period, respectively. P_j is an indicator variable that equals 1 if a CHW *i* participated in the URHVP program at facility *j*. δ_3 and δ_4 are the coefficients of interest for the program and after-program-withdrawal impact, respectively. V_{ijdt} is a set of time-variant CHW characteristics such as the CHW's primary source of income; possession of household assets like agricultural land, livestock, poultry, durables (TV, radio, motorcycle, bicycle, house); and access to financial services (distance to mobile money agent and financial institution). γ_{dt} are district specific linear time trends. μ_i represents CHW fixed effects. These fixed effects

control for the unobserved variables that differ from one CHW to the next, but remain constant over time. ε_{ijdt} is the error term clustered at the district-level. I have clustered at the district level to correct the standard errors for any potential correlation of temporal outcomes for the CHWs within a district.

The identification strategy stated above is based on the common trend assumption. However, my dataset has only one pre-program period (average 2011-2014), which prevents the testing of the parallel trend assumption using multiple pre-program periods. I mitigate this concern by controlling for CHW fixed effects and district specific linear time trends. This should capture any temporal changes in the outcome variables and reduce the estimation bias, if any, arising from the violation of the common trends. Even so, no claim to causality is made.

To ameliorate concerns about false rejections in using many outcome variables, I employ the Romano-Wolf correction method which controls the family-wise error rate (FWER). I include adjusted p-values in the regression results that belong to a family of outcomes (Clarke et al 2019).

3.6 Results and Discussion

3.6.1 Estimation results

The URHVP aimed to encourage the use of skilled maternal healthcare services among poor rural women living in disadvantaged areas. These services would be available during pregnancy, delivery, and in the postnatal period. The project deployed the services of CHWs

and adopted an entrepreneurial model to encourage the workers to deliver on the services under the project. Traditionally, CHWs are volunteers who are selected by their communities specifically to support the provision of primary health care to community members, in addition to various other community activities.

I estimate equation 3.1 to evaluate the impact of the financial incentives provided to CHWs on their service delivery in terms of the outcomes outlined in Section 3.5. Throughout this section, I report the estimates of δ_3 and δ_4 for the program period (post1) and after program withdrawal (post2) impacts, respectively. While I check for an increase in CHWs' efforts for area coverage, and their time use for the provision of incentivized services, I also aim to check if any crowd-out effect is observed in their provision of non-incentivized services, as well as their involvement in community activities due to the introduction of the URHVP.

3.6.1.1 Effect of URHVP incentives on area coverage

I present the results in Table 3.4. Although the results are with and without district specific linear time trends, the main discussion here focuses on results with the influence of district specific linear time trends. In columns 3 and 4, I control for the influence of district specific linear time trends. Based on these results, there is evidence that when compared to the pre-program period, CHW efforts increased (for both program and non-program CHWs) in the program period (indicated in row post1). However, this increase was greater among the program CHWs, by 85.5% and 100%, respectively (indicated in row program x post1 and columns 3 and 4, respectively). Although the program recruited a small number of CHWs to

cover a wide program area, the increase in coverage may be attributed to the financial incentive, because every voucher sold entitled the CHW to a markup of UGX 1,300 (USD 0.35), which was partly used as a transport allowance. If CHW used a bicycle or walked, then all the markup became the CHW's income — a clear motivation to cover more households, but not necessarily more villages, since CHWs started with the nearest villages where they have easy access. Further, since the increased usage of facility services also meant increased revenues for the facility, I cannot rule out the possibility that the health facilities may have strengthened the monitoring and supervision of their CHWs, resulting in increased levels of efforts by the CHW. After the program's withdrawal, efforts returned to the level prior to the program. As indicated, there is no difference between the villages and households covered by the program and non-program CHWs (row program x post2), though it is puzzling to find that the number of villages increased compared to the pre-program period (row post2, column 3). Furthermore, I find significant differences in the coefficients of program x post1 and program x post2 for both villages and households covered (row program x post1 = program x post2, see columns 3 and 4). This is an indication of unsustainable CHW efforts after the program's withdrawal. However, I do not find evidence that suggests the financial incentives backfired on the CHW's moral and efforts.

Altogether, these results indicate that the program increased the service delivery of CHWs in terms of the number of villages and households covered, but only during the program period. After the end of the program, work efforts returned to the pre-program levels, which implies that the temporary incentives may have helped CHWs to overcome their most pressing challenges of lack of transport, and travelling long distances during the

program. However, service delivery may remain stagnant if CHWs are not motivated to overcome these challenges, especially in the rural areas where the accessibility of health facilities is still an obstacle to skilled health care utilization.

3.6.1.2 Effect of URHVP incentives on time use

To supplement the results in 3.6.1.1, I further proxy the effort of CHWs for service delivery by studying the effect of the URHVP on the number of hours per week spent on CHW work and their 1st major economic activity. This was calculated during the program implementation period (post1) and after the program's withdrawal (post2). According to the CHW Strategy of Uganda, CHW work is based on the concept of volunteerism, with no financial incentives except for those provided through implementing partners, as in the case of the URHVP. Therefore, it would be expected that CHWs also attend to other work or economic activities (largely farming for 70% of the CHWs) to sustain their families and livelihoods in the absence of such projects. In this study, I consider the CHW work versus their 1st major economic activity, and present the results in Table 3.5. Similarly, I focus my discussion on the results in columns 7-12 as they take into account the influence of the district specific linear time trends. From the results, there is evidence that during the program period, the program increased the time spent by CHWs (hours per week) on CHW work by 84.1%, which in turn decreased the time dedicated toward their major economic activity by 32.3% (indicated in row program x post1, columns 9 and 12 respectively). In the period after the program's withdrawal, I find no difference between program and non-program CHWs for the time spent on CHW work and the major economic activity (indicated in row program x post2). However, there are significant differences in the coefficients of program x post1 and

program x post2 for both factors (row program x post1 = program x post2, columns 9 and 12). This is an indication that the withdrawal of the incentive leads CHWs to reduce their time for CHW work and substitute it with their economic activity. As with the major economic activity, the financial incentives are clearly considered as a source of livelihood. This is why CHWs diverted their time to social work for the incentive, and reverted to their economic activity when the incentive expired.

In sum, the program increased the CHWs' working hours per week but decreased the hours they spent on their other economic activities. This switching trend between CHW work and the economic activity during and after program implementation suggests the need to integrate financial incentives into the CHW strategy as means to sustain the CHW's performance.

3.6.1.3 Effect on provision of incentivized and non-incentivized services

The main tasks for CHWs under the URHVP included the following: sensitization of pregnant women, selling of vouchers, follow-up of women to use the voucher services (antenatal care, delivery and postnatal care), follow-up of women to use the non-voucher services (immunization of newborn babies), and submission of reports on the program. All the services related to these tasks are considered as incentivized services, while the others are non-incentivized services. I have already demonstrated that financial incentives improved the area of coverage and the weekly working hours devoted to CHW work. This naturally raises the next question — does this increase also influence the provision of incentivized services and does it crowd-out the non-incentivized services in exchange for program related

services, all of which may negatively affect people with health problems other than maternal health? In the same vein, I also examine if the CHWs who were provided financial incentive felt discouraged after the incentive was ceased. Some programs using the entrepreneurial model may be successful during the program period, but can change the work ethic of the CHWs (those who were socially motivated before the program, but whose mentality on "volunteer work" is no longer the same), resulting in negative unintended consequences.

Table 3.6 presents the regression results on the impact of the URHVP on CHW efforts toward incentivized (in columns 1-6 and 9-14) and non-incentivized services (in columns 7-8 and 15-16), measured by the number of days per month spent on the provision of these services. To account for differences in district characteristics and the possibility that the provision of services evolved differently in facilities that received the program compared with those that did not, I included district specific linear time trends in columns 9-16. Thus, these results are the basis for my discussion. From the previous results, I found a large increase in the number of households and villages covered (Table 3.4 columns 3 and 4) as well as the time spent on CHW work (Table 3.5 column 7 shows a 46% increase in the days per week spent on CHW work). However, for most of the incentivized services, it is puzzling to note that there is no evidence that the program increased the CHW's efforts (number of days per month) for the provision of these services during and after the program, as there is no difference between the program and non-program CHWs (indicated in rows program x post1 and program x post2, respectively for columns 10-14). Although the effort levels reverted to the pre-program levels, there is evidence of increased CHW effort in health education during the program period. This is economically significant at 37.3% in column 9

(indicated by the interaction term program x post1). This significance may exist because health education was the CHWs core task (to provide information about the voucher, and to motivate eligible mothers to buy the voucher and seek skilled maternal care services).

In the program period, the CHW efforts towards data collection & reporting, and other activities decreased compared to the pre-program period (row post1, columns 12 and 14, respectively). Furthermore, the CHW efforts towards health education, health campaigns, case referrals, data collection, and other activities in the post-program period declined compared to the pre-program period (row post2, columns 9, 10, 11, 12 and 14, respectively). However, there are no differences between the program and non-program CHWs. Having said that, I note significant differences in the coefficients of program x post1 and program x post2 for health education and women counselling services (row program x post1 = program x post2, columns 9 and 13). This is an indication of unsustainable CHW efforts toward the provision of these services after the program's withdrawal.

Despite the increased CHW service delivery (in terms of coverage) and working hours per week by the program (Tables 3.4 and 3.5 above), there does not seem to be crowding out of the non-incentivized services, such as treatment of non-complicated diseases, and the provision of HIV/AIDS preventive education and counselling services (indicated in rows program x post1 and program x post2, columns 15 and 16, respectively). Although the program did not crowd out non-incentivized services provided by the CHWs, the withdrawal of incentives leads to a decline (albeit small) in the efforts of CHWs toward the provision of HIV/AIDS services (as indicated in program x post1=program x post2, column 16).

In summary, while I found large increases in the area of coverage and CHW time spent on service provision, it is puzzling that the program did not have much of an effect on the incentivized services. This might be because the program only increased CHW efforts in the provision of health education services, and this too reverted to the pre-program levels by the end. Additionally, there was an unsustainable provision of women counselling services after the program withdrawal. For the non-incentivized services, the increased CHW service delivery (coverage) and working hours per week does not seem to have resulted in their being crowded out during the program period, though the provision of HIV/AIDS related services did become unsustainable after the program. By and large, the unsustainable effects on the efforts of the CHWs is an indication that their work ethic is motivated by financial incentives in one way or another.

In the study by Wagner et al. (2019), CHWs working under the home sales arm for dispensing ORS and Zinc earned about UGX 16,120 (USD 4.38) per month while under the URHVP, excluding the transport allowance from voucher sales, a CHW earned about UGX 30,000 (USD 8.16) per month in redemption bonuses. In this case, the profit made by the program CHWs under the URHVP is higher, and one would expect this incentive to motivate program CHWs to deliver exceptional services. These results may also imply that despite the financial incentives provided, the challenges faced by the CHWs, such as the lack of transport, travelling long distances, and the lack of support from government officials (as indicated in Table 3A3), remain a major constraint for service delivery. Additionally, the motivation of CHWs may also be influenced, albeit to a small extent, by their expectations as indicated in

Table 3A4. These include their expectations of financial gains, employment opportunities and political support.

3.6.1.4 Effect of URHVP incentives on involvement in prosocial community activities

Although the CHWs provide awareness and generate demand for health services, they also participate and lead their communities in other cross cutting activities such as charity work, volunteering, and political party activities. For these activities, respondents were queried on how many times they got involved or participated before the program, during the program's implementation (post1), and after the program's withdrawal (post2). I defined three dummy variables for charity, volunteering, and political party work, each taking the value of 1 if a CHW was involved in an activity at least more than once. Thus, since the URHVP incentivized CHWs to work more, their involvement in other community work (charity, volunteer, and political) should have decreased (during the program implementation), which may not be a good thing for their communities. Using equation 3.1, I also examined if such a crowding out effect is observed as a result of the introduction of URHVP. I present the results in Table 3.7 with columns 4-6 controlling for the influence of district specific linear time trends.

To examine the crowding out effect on social activities as a result of the introduction of the URHVP, I would expect to find a negative (and significant) coefficient during the program and a positive (and/insignificant) coefficient after the program withdrawal. However, the results do not show this pattern consistently. The results suggest that there is no evidence that participation in the program increased or decreased CHWs' involvement in

community prosocial activities (charity, volunteer, and political party work) both during and after the program (indicated in rows program x post1 and program x post2, respectively). In comparison with the pre-program period, there was an increase in the program and non-program CHWs' involvement in charity (by 23.8 percentage points), volunteering (by 36.1 percentage points) and political party (by 20.9 percentage points) activities during the program period (indicated in row post1, columns 4, 5 and 6, respectively), though there is no difference between the program and non-program CHWs (as indicated in row program x post1). The general increase in the involvement of CHWs in these prosocial activities may be attributed to the general elections that took place in Uganda in 2016. In the period following the program's withdrawal, I find evidence of increased involvement of program and non-program CHWs in volunteer work (by 50.2 percentage points) as compared to the pre-program period (row post2, column 5). However, this increase was greater among the program CHWs — by 21.6 percentage points (row program x post2, column 5). Furthermore, since there is no significant difference in the coefficients of program x post1 and program x post2 for all prosocial activities, this shows a sustainable level of involvement by CHWs in these activities even after the program was withdrawn.

In summary, I find no evidence for the crowding out of community activities for CHWs due to the introduction of the URHVP — a good thing for the communities in which the CHWs serve. These results are in contrast to Dal Bo et al. (2013)'s results, who found that financial incentives had no effects on volunteer work, while they did negatively impact charity work and political party participation. Despite this, these results should be interpreted cautiously. As my dummy variable equals 1 if a CHW was involved in an activity more than

once, I cannot tell whether there was an actual increase or decrease in involvement or time spent on community work. The program period coincided with an election, as a result of which I see an increased likelihood of CHWs being involved in political party activities. However, this does not mean that those who were involved in such activities increased the number of days/hours of their involvement.

Therefore, during the program period, the program increased service delivery in terms of villages and households covered by CHWs, and increased their working hours per week, but decreased the weekly hours spent on other economic activities. Furthermore, these increases by the program do not seem to be realized by the crowding out of non-incentivized services. The effects on CHW coverage, time use, and the provision of services are unsustainable because the CHW effort levels reverted to the pre-program levels when the incentive was withdrawn. All the above results are not different from the sub-sample findings (not shown for brevity). Further, I found no evidence that participation in the program increased or reduced participation in other social activities both during and after the program. However, these results should be interpreted with caution, as discussed earlier.

3.6.2 HETEROGENOUS EFFECTS

So far, my analyses have assumed that the URHVP had a homogenous impact across all the CHWs of different characteristics. However, it is possible that the impact of the program on the CHW's effort may vary based on multiple factors including: 1) CHW's gender, 2) how the CHW was selected (by community popular vote or otherwise) and 3) economic status of the CHW.

First, one may wonder if the CHW's gender changes the impact of the maternal program. For example, do female CHWs have better networks among mothers and find it easier to gain access to pregnant women than their male counterparts, or could male CHWs have greater service delivery as they visit more villages and households simply because they are more flexible and able to travel longer distances than women? I defined a dummy variable that equals 1 if the CHW was female, and zero if the CHW was male ($female_v$).

Secondly, I also investigated heterogeneity based on the selection of the CHWs. CHW work involves and requires community participation; most of which is done by visiting community members in their villages and households. CHW respondents were asked how they were selected for this role.[46] I defined a dummy variable that equals 1 if the CHW was selected through community popular vote, and 0 if selected by the district/community leaders, health officials, and/or the program development/implementing partners ($popular_v$).[47] I would expect the appointed CHWs (not selected through the community popular vote) to face some resentment from the community, which would negatively affect their service delivery in the community.

[46] How were you selected as a VHT? 1) by district chairperson; 2) by district health officers; 3) through community popular vote; 4) through a development program; 5) by local leaders; 6) by myself; and 99) other.

[47] The share of CHWs selected based on community popular votes does not vary across districts or across facilities. On average it is 70%. In program districts – Community popular votes 76.43%, others 23.57%; In non-program districts – Community popular votes 70.45%, others 29.55%; At program facilities – Community popular votes 76.29%, others 23.71%; At non-program facilities – Community popular votes 72%, others 28%.

Lastly, CHW asset ownership provides a prediction of the economic status of the CHW. I expect the impact of the program incentives to be larger among the poorer CHWs (those who are initially worse-off) because they would put in greater effort on service delivery to earn more from the program. I constructed the asset index and quintiles as detailed in Appendix 3B. I considered the bottom 40% (quintile 1 and 2) as poor, equaling one, and zero if otherwise ($poor_v$), expecting that the effect of the program may be larger among the poor CHWs than those that were better-off.

I therefore estimate the heterogenous treatment effects for the CHW's gender, selection process, and economic status by adding the respective interaction terms to the model in equation 3.2 below, for both the full and sub sample:

(3.2) $Y_{ijdt} = \delta_0 + \delta_1 Post1_t + \delta_2 Post2_t + \delta_3(\phi_i \times Post1_t) + \delta_4(\phi_i \times Post2_t) + \delta_5(P_j \times Post1_t) + \delta_6(P_j \times Post2_t) + \delta_7(P_j \times Post1_t \times \phi_i) + \delta_8(P_j \times Post2_t \times \phi_i) + V_{ijdt} + \gamma_{dt} + \mu_i + \varepsilon_{ijdt}$,

where ϕ_i represents the dummies for $female_v$, $popular_v$ and $poor_v$. Coefficient δ_7 (δ_8) captures the differential treatment effect of the program on CHWs (female, popular and poor) compared to CHWs (male, appointed and non-poor) during the program period (and after the program's withdrawal).

Table 3A5 (CHW effort by coverage), Table 3A8 (CHW effort by time use), Table 3A11 (incentivized and non-incentivized services) and Table 3A14 (community activities) present the results for the heterogenous impact by gender. In all results, the coefficients of

the interaction terms (Female x Program x Post1 and Female x Program x Post2) are insignificant, suggesting no evidence exists that the effect of the program is different between female and male CHWs.

In Table 3A6 (CHW effort by coverage), Table 3A9 (CHW effort by time use), Table 3A12 (incentivized and non-incentivized services) and Table 3A15 (community activities), I show the results for the heterogenous impact based on the CHW selection (selected through community popular vote). Once again, there is no difference between the CHWs selected through the popular vote and those appointed as CHWs in the effect of the program on their weekly hours for CHW work and economic activities (Table 3A9), as well as their involvement in the community activities of charity, volunteering, and political party activities (Table 3A15). This is because the coefficients of the interaction term ($Popular_v$ x Program x Post1 and $Popular_v$ x Program x Post2) are both insignificant. Meanwhile, I find that the increase in the number of households covered by the CHWs (Table 3A6, row $Popular_v$ x Program x Post1 in full sample) and the increase in the effort for the provision of case referral services, data collection and reporting, and HIV/AIDS services (Table 3A12, rows $Popular_v$ x Program x Post1 and $Popular_v$ x Program x Post2) are more concentrated among the popular CHWs.

In Table 3A7 (CHW effort by coverage), Table 3A10 (CHW effort by time use), Table 3A13 (incentivized and non-incentivized services) and Table 3A16 (community activities), I present the results for heterogenous impact based on the economic status of the CHWs (CHW asset index of bottom 40% as poor). In all results, the coefficients of the

interaction term ($Poor_v$ x Program x Post1 and $Poor_v$ x Program x Post2) are insignificant, suggesting no evidence that the effect of the program is different between poor and non-poor CHWs, except for the reduction in efforts made for health campaigns and women counselling services. This was common among the poor CHWs in the period after the program's withdrawal.

In sum, these heterogeneity analyses show no evidence that the effects of the incentives were different by gender, both during and after the program, while the effects on household coverage and service provision (case referrals, data collection, and HIV/AIDS related services) were greater for CHWs who were selected by the popular vote than those who were not. Further, I also find that in comparison to the better-off CHWs, the program negatively affected the motivation of the poorer CHWs to a greater extent by reducing their efforts for the provision of health campaigns and women counselling services, especially after the withdrawal of the incentives.

3.6.3 ROBUSTNESS CHECKS

I have conducted a number of different robustness checks on my results. First, in Table 3.1, I found that the program CHWs are more likely to be married/living with a partner and have completed at least primary level education than the non-program CHWs. Since these two characteristics are significantly different between the groups of CHWs at the program and non-program facilities, I constructed a sub-sample maintaining only those CHWs that are married/living with a partner, and have completed at least primary education. Second,

since the URHVP survey of CHWs is not nationally representative, I obtained sampling weights to account for the over-sampling of program facilities.[48] Based on this, I re-estimated the models for the full and sub-samples including the sampling weights. The results for both sets of samples are presented in Tables 3A17 – 3A20. The results reported for both samples are very close in magnitude and statistical significance to the main results in Tables 3.4 – 3.7.

Third, there is a concern that the CHW time-variant characteristics may drive the average effect of the program. To counter this possibility, I re-estimated the models for the full sample by dropping all time-variant characteristics. The results reported in Tables 3A21 and 3A22 are very close in magnitude and statistical significance to the main results in Tables 3.4 and 3.5. Further, the results in Table 3A15, on the effort of CHWs for the provision of non-incentivized services and the sustainability of this effort after the program's withdrawal are close to the earlier results in Table 3.6. However, unlike the main results in Table 3.6, the results reported in Table 3A23 indicate that the large increase in the number of households and villages covered as well as the hours spent on CHW work, led to a much larger effect on the provision of all incentivized services during the program period. Lastly, the results on the likelihood of CHWs' involvement in community activities shown in Table 3A24 are almost similar to the earlier results in Table 3.7, except for the increased likelihood of involvement in volunteer activities by the program CHWs during the program (extensive margin).

[48] I compute the district-level weight for non-program facilities and program facilities as:

District-level weight for districts for non-program facilities = $\frac{total\ number\ of\ non-program\ facilities\ in\ districts}{total\ number\ of\ sampled\ non-program\ facilities\ in\ districts}$

District-level weight for districts for program facilities = $\frac{total\ number\ of\ program\ facilities\ in\ districts}{total\ number\ of\ sampled\ program\ facilities\ in\ districts}$

However, these should be interpreted with caution because it does not mean that those who were involved in such activities increased the number of days/hours of their involvement (changes on the intensive margin).

Fourth, I rely on the CHW survey responses for most of our outcomes as I did not have access to any reports at the district health offices. One may argue that CHWs overstate their service delivery. To address this concern, I analyze the data from the health facilities survey using equation 3.3:

$$(3.3) \qquad Y_{jdt} = \delta_0 + \delta_1 Post1_t + \delta_2 Post2_t + \delta_3(P_j \times Post1_t) + \delta_4(P_j \times Post2_t) + \gamma_{dt} + \varphi_j + \varepsilon_{jdt},$$

where Y_{jdt} indicates the outcome of interest at health facility j in district d in period t — an indicator variable that equals 1 if CHWs participate in providing services at the health facility, and zero otherwise. $Post1_t$ and $Post2_t$ are indicator variables that equal 1 for the program-implementation period and the post-program period, respectively. P_j is the program indicator variable that equals 1 if a health facility j was supported by the URHVP. γ_{jdt} are district specific linear time trends, while φ_j are health facility fixed effects.

Table 3A25 presents the results of the perceptions of health officials on the participation of CHWs' and health care service provision. In columns 5-8, I control for the influence of district specific linear time trends. Based on these results, the program increased the CHWs' attendance of community health meetings and training at the program facilities by 24.6 percentage points (indicated in rows program x post1, column 8). However, when compared

to the pre-program period, there is also evidence of decreased CHWs' community visits (for both program and non-program facilities) during and after the program (indicated in rows post1 and post2). Yet, there is no difference between the program and non-program facilities (indicated in rows program x post1 and program x post2, column 6). Although I find no difference in the program's impact on CHWs' participation in providing services at program and non-program facilities during the program period (row program x post1), the withdrawal of incentives led to a decline (albeit small) in the participation of CHWs across all services (as indicated in program x post1 = program x post2, columns 5 to 8). The results, particularly on the community visits by CHWs may be inconsistent with the results seen in Table 3.4. This is either an indication of over-statement by the CHWs, or the health facility respondents may not be fully aware of what CHWs do.

Further, there may be the possibility that the increased service delivery by CHWs resulted in the deterioration of quality. Using the survey conducted by the health facility and women who used the health facility, I test if the program affected the quality of CHWs' services during and after the program. Table 3A26 indicates that the quality of services provided by the CHWs did improve at the health facilities during the program.

3.7 CONCLUSION

This study set out to examine the impact of the URHVP on CHWs' efforts toward service delivery and involvement in community work. More specifically, I investigated if the short-term improvements in service delivery are driven by the temporary financial incentives

to CHWs and if the services and community activities not considered for URHVP payments are crowded-out in exchange for program related services. The latter is an important issue since such incentives may affect community members with health concerns other than maternal health, and the community at large. In addition, I also investigated if CHWs who were provided financial incentives were discouraged after the incentive was no longer provided. This is because it could affect their work ethic and result in negative unintended consequences.

To examine these issues, I utilized the panel data collected from CHWs at the URHVP program and non-program facilities. Applying the difference-in-differences approach with CHW fixed effects, I show that the program increased service delivery in terms of the villages and households covered (by 85.5% and 100%, respectively) and increased the working hours per week (by 84.1%), which was realized by decreasing the weekly hours spent on other economic activities (by 32.2%). However, there is no evidence that the program discouraged CHWs from providing non-incentivized services and decreased the likelihood of their participation in community work. Contrary to Ashraf et al. (2014) and Wagner et al. (2019), these results indicate that financial incentives are a powerful mechanism to enhance CHW efforts, especially for the period that they last.

The estimated positive effects of the program on CHWs' service delivery during the program period are found to be unsustainable as their effort level reverts to the pre-program levels once the program is withdrawn. Unlike Celhay et al. (2019), where the effects of incentives for the adoption of better-quality prenatal care practices persisted for at least 24

months after the incentives ended, the effect of incentives on CHW effort in my study did not persist after the withdrawal of the incentives. The key difference is that in Celhay et al. (2019), the motivation for the provision of incentives was the adoption of improved strategies and better quality practices at the clinics. This impact was likely to last longer, which was not the case in our study where the incentives were provided to CHWs, who typically do not receive any financial reward from the government. It is no wonder then that when the incentives expire, their motivation decreases because they may not be able to easily move around in villages and to households to provide services due to the lack of transport and other related challenges.

I also add to the existing body of literature by extending my analysis to investigate if there are differential treatment effects among the CHWs. The heterogeneity analyses show no evidence that the effects of the incentives were different by gender both during the program period and after the program's withdrawal. However, the effects on household coverage and service provision (case referrals, data collection, and HIV/AIDS preventive education and counselling) were more pronounced for the CHWs who were selected by community popular votes than those who were not. Furthermore, the program negatively affected the motivation of the poorer CHWs more by lowering their efforts toward health campaigns and women counselling services once the incentives were withdrawn.

Some limitations should be considered when interpreting these results. First, the survey sought retrospective data, before the program (period 2011-2014), during the program (period 2015-2019) and after program withdrawal (6 months period 2019-2020). Although

these periods were distinctive and respondents could easily remember the data requested, I acknowledge the possibility of recall error in the responses to retrospective questions. Second, I cannot generalize my findings to all CHWs because this study sample is not representative of all CHWs in Uganda, especially since the URHVP selected the active ones only. Therefore, the impacts ascertained are when active CHWs were selected to the program. There may be a two-way direction of any possible bias for the impacts. If active CHWs work hard even without financial incentives and/or if financial incentives are likely to discourage their effort, then the marginal effect of the program is likely to be smaller than that when selecting CHWs randomly. By contrast, if active CHWs work even harder with financial incentives than the less active CHWs, then it is also likely that the impacts are underestimated when only using active CHWs compared with using randomly selected CHWs. Third, I found large increase in area coverage and CHW time for service provision, although the effect on provision of incentivized services is not much, except for the increased effort in provision of health education services. This brings to question, what the increased coverage and hours worked were used for. In addition, as the program comprised of several components, one may query if the impact on CHW effort and work ethic is a result of all program components. However, although factors like training and supervision among others, may contribute to the observed program effects, I cannot disentangle the effect of incentives from such components. Fourth, I find no evidence that the program crowds out the involvement of CHWs in community activities. The measure of involvement in community activities is a dummy variable if one was involved in such activities during the last 12 months. It captures the change only when one stops getting involved in such activities (extensive margin). It does

not change even if one decreases their involvement in these activities (intensive margin). For the future, this effect may be analyzed by considering the time of involvement (hours or days) in these activities instead. Fifth, unlike the study by Celhay et al. (2019), this survey was conducted 6 months after the closure of the program and my measurements for the program withdrawal period consider this 6-month period alone. Since the CHWs may require some adjustments after the program is withdrawn, the estimated impacts in this study might be underestimated. To examine the long-term sustainability of the program, one might argue that this 6-month period is too short to provide a true reflection of the program's long-term impact. Therefore, there is a need for future analysis of this effect after a longer period. Lastly, I acknowledge the possible bias of using DID for treatment implementation at the multiple time periods as pointed out by Callaway and Sant'Anna (2021) and others. I also acknowledge that the average data collected for "during program period" under the CHW survey has a limitation as the year of actual start of the program differed across program facilities ($Post1_t$ does not refer to the same calendar periods across facilities). My future revisions will take these concerns into account.

CHAPTER 4

CONCLUSIONS AND POLICY IMPLICATIONS

4.1 CONCLUSION

The level of access, utilization, and quality of skilled maternal health care services is low among the rural poor in developing countries. This is due to multiple barriers. Various governments have adopted a variety of policies to increase the utilization of maternal care while taking into account constraints on the supply and demand sides. One such policy is the integration of maternal voucher programs into the health care system. Maternal voucher programs have been widely adopted to promote maternal and child health in low and middle-income countries for the last 15 years at least (Hunter BM. et al., 2017). On the supply side, these programs aim to improve the quality and responsiveness of service provision, while on the demand side, they aim to encourage the utilization of services by reducing the financial burden, and educating beneficiaries on the benefits of utilizing skilled maternal health care services. While voucher programs have successfully improved the usage of maternal care services so far (Nguyen et al., 2012; Van de Poel et al., 2014; Mohanan M. et al., 2014: Alfonso et al., 2015), there are some concerns regarding these programs. The first concern is that when the programs finish, it is not clear if the situation will revert to worse than before the program. This is because fiscal constraints do not allow governments to afford a subsidy forever. Second, though the usage of maternal care services may increase, does the quality-of-service provision match the demand for services, and does this lead to the desired and sustainable health outcomes? Third, though vouchers target rural poor women and the service providers in their proximity, are the positive effects larger in areas with poorer households.

Chapter 2 of this dissertation evaluates the Uganda Reproductive Health Voucher Program (URHVP) to explore these concerns, both during the program implementation and after the program's withdrawal.

Meanwhile, in order to improve the utilization of healthcare, provision of services, and health outcomes, the use of Community Health Workers (CHWs), a community-based strategy, has also been widely adopted and integrated in the health care systems and programs (Deserranno, 2019). Although largely based on the concept of volunteerism, there is growing support of entrepreneurial and financial incentives in CHW programs to motivate the efforts and performance of CHWs (Basinga et al., 2011; Miller et al., 2012; Finan et al. 2017; Nyqvist et al., 2019). By contrast, some evidence suggests that CHWs perform better when they are non-financially motivated because they are socially motivated by nature and financial rewards can distort their inherent motivation (Ashraf et al., 2014; Wagner, Z. et al., 2019). Some concerns have remained unanswered in the existing literature when estimating if financially or socially motivated CHWs are more effective at improving the provision healthcare services. In Chapter 3 of this dissertation, I evaluate the same maternal voucher program which provided incentives to the participating CHWs (who previously did not receive any financial rewards from the government). I not only seek to establish its impact on the efforts of CHWs (by coverage and time spent on CHW work), but also, to determine if these program incentives crowd-out efforts toward other non-incentivized services and involvement in community services. I also explore if the CHWs' mentality on volunteer work remains the same when the incentive is no longer provided.

4.2 SUMMARY OF FINDINGS

In Chapter 2 of this dissertation, I investigated how the URHVP support affected the use of maternal care, quality of maternal care services, and health outcomes for mothers and children during and after the program. I applied the difference-in-differences approach with i) health facility fixed effects to the annual administrative health facility MoH HMIS data and health facility survey data; and ii) women fixed effects to the survey data of women who used the health facilities. The results show that during the program period, the maternal voucher program enhanced the usage of all maternal care services (ANC (4^{th} visit), delivery, and PNC); the quality of ANC, delivery care, and PNC services; and the number of doctors, nurses, and midwives during the program implementation period. However, after the program's withdrawal, both usage and quality levels reverted back to the pre-program levels as there was no evidence for any distinct effect. There was also no evidence that the URHVP improved child health outcomes, such as infant birthweight, reduced complications during delivery, and improved pregnancy outcomes, such as increased livebirths, or reduced stillbirths, miscarriages, and neonatal deaths. Lastly, my investigation on whether the treatment effects were greater in poorer districts than in better-off ones indicated pro-poor effects on the number of medical doctors in the health facility after the program was withdrawn. However, these results yield no evidence that the effect of the program on most of the outcome variables is larger in the poorer districts. This is an indication that the program had same positive impacts. Overall, the results support the implementation of voucher programs to improve the utilization and quality of care, though there is an urgent need to solve issues related to program sustainability and the improvement of health outcomes.

In Chapter 3, using a similar approach, I applied the difference-in-differences method with CHW fixed effects to the original URHVP CHW survey data to establish the effect of the program incentives on CHWs' effort and work ethic, and examine if the CHWs traded non-incentivized services for program services during and after the program period. The results show that the program increased service delivery in terms of the villages and households covered and increased the working hours per week, which was realized by decreasing the weekly hours CHWs spent on other economic activities. Further, there is no evidence that the program caused a decrease in the provision of non-incentivized services and the likelihood of CHWs participating in community work. However, the estimated positive effects of the program on CHWs' service delivery during the program period were unsustainable, as the effort of CHWs reverted to the pre-program levels when incentives were withdrawn. Overall, I note that financial incentives are a powerful mechanism to enhance the performance of CHWs, at least for the period that they last. What needs to be addressed is a solution to sustain CHWs' performance gains beyond the programs that support them.

4.3 Policy Implications

The findings from this study have substantial policy implications as they point to the vital role comprehensive voucher programs can play in i) promoting maternal health care utilization, ii) improving the performance of service providers and community agents in the provision and quality of health care, and as a result, iii) improving maternal and child health outcomes. They show that governmental and donor commitment can significantly improve health care systems in developing countries, where fiscal constraints are prevalent. The study

findings suggest that complementary voucher programs (which not only provide vouchers to women, but also give supply-side monetary incentives to health care providers and agents) can improve the usage and quality of healthcare services at health facilities, at least for the duration of the program. Unfortunately, they failed to achieve post-program sustainability and a significance improvement in child and maternal health outcomes. This calls for policy makers to devise means of preventing reversion to the pre-program effects. For example, supply side interventions (through public-private partnerships) to keep the cost of services low, while maintaining service quality, could help maintain the above-mentioned positive program effects. Since, healthcare as a component of the social transformation aspiration in the country's Vision 2040, proposes for a shift from a public centered health delivery system to a public-private-partnership, as one of its strategies for the sector.[49] Nsasira et al., (2013), argue that a public-private partnership can be a sustainable funding mechanism where legal, technical and administrative capacity is developed.

Additionally, observing that CHW efforts return to pre-program levels when incentives expire also brings into question what policy makers ought to do to sustain the performance of CHWs. The financial incentives paid to CHWs were aimed to boost their efforts in the delivery of program services, and to reduce the challenges of low motivation and lack of transportation to make visits (both being long-standing challenges). The volunteerism concept alone may not sustain CHW performance because some workers may not be in position to trek long distances. It is no surprise then that they trade off their

[49] Uganda Vision 2040, 2017 accessed online on June 16, 2021, from http://library.health.go.ug/publications/leadership-and-governance/uganda-vision-2040

economic activities for CHW work when a financial opportunity is presented, and the reverse is true when the financial opportunity expires. Therefore, integrating modest financial rewards into the CHW strategy may be effective in delivering sustainable performance. One might consider that low-income governments cannot afford permanent funding for community level activities such as those by CHWs. However, without underestimating the financial constraints of 3rd world countries, in the case of Uganda, there are still multiple efforts to uplift households from poverty conducted at the household and community levels that have been sustainably funded by the government and/or overseas development assistance over the decades. As examples, the focus has lately been on initiatives like Wealth Creation and Job creation (EMYOOGA), and Parish Development model (PDM), among others. The work of CHWs can be incorporated into this development framework with measurable deliverables (related to the project outcomes), especially given that CHWs play a larger role beyond just health related community support. Therefore, as health is a critical dimension of the existing framework for household and community level development (Uganda Vision 2040)[50], funding to these social levels can also target CHWs as the vehicles of sustainable development.

Overall, in order to sustain the positive effect of such programs, it is crucial that the program designers and implementers focus on program sustainability and significant improvements in child and maternal health outcomes to avoid stagnation in the healthcare system.

[50] Uganda Vision 2040, 2017 accessed online on June 16, 2021, from http://library.health.go.ug/publications/leadership-and-governance/uganda-vision-2040

4.4 FUTURE RESEARCH

This dissertation examines the role of a comprehensive maternal voucher program in improving the use of skilled maternal healthcare services among poor rural women living in disadvantaged areas, and the impact of the program on improving the quality of services provided at the health facilities. In addition, I examine the program's effects on CHW efforts, their work ethic after program expiry, and if any crowding-out of non-incentivized services and community activities is found as a result of the introduction of the program. It is hoped that these findings will motivate further empirical studies on the issues of program sustainability.

In Chapter 2, the women sample was determined by the CHWs and only contained women who had recently delivered multiple children, or had multiple pregnancies. This poses selection issues and may limit the external validity of the study's findings. Future analysis may consider analyzing the effect on maternal care utilization and child and maternal outcomes using randomly selected mothers. Second, I could not examine if the effect of the program is greater for poorer women than for better-off women because we did not collect income data from the pregnant women surveys. Since the program specifically targeted poor pregnant women incapable of meeting the out-of-pocket expenses for maternal care, the heterogenous effects of the program by poverty status should be investigated using household-level income. Third, in the context of the long-term sustainability of the program, there is a need for future analyses of the program's effects on the use and quality of maternal care after a longer period. Fourth, despite being used by other studies, I relied on the women's self-reported measures to assess the maternal and child health outcomes. Future studies that

examine other indicators based on their medical records may be more fruitful. Lastly, as the year of actual start of the program differed across program facilities, I acknowledge that the average data collected for "during program period" under the Health Facility survey has some limitation ($Post1_t$ does not refer to the same calendar periods across facilities). My future revisions will require that I devise some ways to control for this. In addition, future revisions will also take into account the possible bias of using DID for treatment implementation at the multiple time periods as pointed out by Callaway and Sant'Anna (2021) and others.

In Chapter 3, as my findings are based on the most active CHWs attached to the sample health facilities, they may not be generalizable to all CHWs. I therefore suggest using randomly selected CHWs to analyze the effect on their efforts. Second, although, I found large increase in area coverage and CHW time for service provision, it is puzzling that the effect on provision of incentivized services is not much, except for the increased effort in provision of health education services. In addition, as the program comprised of several components, one may query if the impact on CHW effort and work ethic is a result of all program components. However, potentially contributing factors like training and supervision among others, may contribute to the observed program effects but I cannot disentangle the effect of incentives from such components. Future studies that shed more light on what the increased coverage and hours worked were used for and possible disintegration of other contributing factors from incentives, may be fruitful. Third, with respect to CHW involvement in community activities, I conclude that there is no crowding-out effect. This may be due to their increased participation in community activities (extensive margin), but I also acknowledge the possibility that there could have been some changes for prosocial

behavior on the intensive margin. Therefore, I suggest future analyses of this effect by considering how frequently or for how many days/hours a CHW participates in these activities. Fourth, for the period after program withdrawal, I considered 6 months since the survey was conducted 6 months after the closure of the program. Therefore, my estimated impacts may be underestimated and not a true reflection of the long-term impact of the program. Thus, there is a need for future analyses to shed more light on the long-term post voucher program trends. Lastly, the CHW survey data collected for "during URHVP program period" does not take into account the varying program intensity across program facilities ($Post1_t$ does not refer to the same calendar periods across facilities). Future analysis may require that I devise some ways of controlling for this. In addition, future revisions will also take into account the possible bias of using DID for treatment implementation at the multiple time periods as pointed out by Callaway and Sant'Anna (2021) and others.

REFERENCES

Abdi, H., 2007. Bonferroni and Šidák corrections for multiple comparisons. Encyclopedia of measurement and statistics, 3, 103-107.

Agha, S. (2011a). Changes in the proportion of facility-based deliveries and related maternal health services among the poor in rural Jhang, Pakistan: Results from a demand-side financing intervention. *International journal of equity health*, 10.

Agha, S. (2011b). Impact of a maternal health voucher scheme on institutional delivery among low income women in Pakistan. *Reproductive health, 8*(1), 10.

Alfonso, Y. N., Bishai, D., Bua, J., Mutebi, A., Mayora, C., & Ekirapa-Kiracho, E. (2013). Cost-effectiveness analysis of a voucher scheme combined with obstetrical quality improvements: quasi experimental results from Uganda. *Health policy and planning*, *30*(1), 88-99.

Ashraf N., Oriana B. & Kelsey Jack B. (2014). No margin, no mission? A filed experiment on incentives for public service delivery. *Journal of Public Economics* 120: 1-17.

Ashraf, Nava; Bandiera, Oriana; Davenport, Edward; Lee, Scott S. (2020)."Losing Prosociality in the Quest for Talent? Sorting, Selection, and Productivity in the Delivery of Public Services." *American Economic Review*, 110 (5): 1355-94. doi: 10.1257/aer.20180326

Banerjee, A. V., Duflo, E., & Glennerster, R. (2008). Putting a Band-Aid on a Corpse: Incentives for Nurses in the Indian Public Health Care System. *Journal of the European Economic Association*, 6(2-3), 487–500. doi:10.1162/jeea.2008.6.2-3.487

Basinga, P., Gertler, P. J., Binagwaho, A., Soucat, A. L., Sturdy, J., & Vermeersch, C. M. (2011). Effect on maternal and child health services in Rwanda of payment to primary health-care providers for performance: an impact evaluation. *The Lancet*, 377(9775), 1421–1428. doi:10.1016/s0140-6736(11)60177-3

Blum, Florian (2020). The value of discretion: Price-caps and public service delivery. *Journal of Development Economics*, 146(), 102521. doi:10.1016/j.jdeveco.2020.102521

Callaway, B., Sant'Anna, P. H. C., December 2021. Difference-in-differences with multiple time periods. *Journal of Econometrics* 225 (2), 200–230.

Celhay, Pablo A., Paul J. Gertler, Paula Giovagnoli, and Christel Vermeersch. 2019. "Long-Run Effects of Temporary Incentives on Medical Care Productivity." *American Economic Journal: Applied Economics*, 11 (3): 92-127.

Clarke, D, Romano J.P. and Wolf M., (2019). The Romano-Wolf Multiple Hypothesis Correction in Stata, Forthcoming, Stata Journal.

Cohen, J., & Dupas, P. (2010). Free Distribution or Cost-Sharing? Evidence from a Randomized Malaria Prevention Experiment*. *Quarterly Journal of Economics*, *125(1), 1–45*. doi:10.1162/qjec.2010.125.1.1

Dal Bo, E.; Finan, F.; Rossi, M. A. (2013). Strengthening State Capabilities: The Role of Financial Incentives in the Call to Public Service. *The Quarterly Journal of Economics*, 128(3), 1169–1218. doi:10.1093/qje/qjt008

Dennis, M. L., Abuya, T., Campbell, O. M. R., Benova, L., Baschieri, A., Quartagno, M., & Bellows, B. (2018). Evaluating the impact of a maternal health voucher programme

on service use before and after the introduction of free maternity services in Kenya: a quasi-experimental study. *BMJ global health*, *3*(2), e000726.

Deserranno, Erika. 2019. "Financial Incentives as Signals: Experimental Evidence from the Recruitment of Village Promoters in Uganda." *American Economic Journal: Applied Economics*, 11 (1): 277-317.DOI: 10.1257/app.20170670

Duflo, E., Hanna, R., & Ryan, S. P. (2012). Incentives Work: Getting Teachers to Come to School. *American Economic Review*, 102(4), 1241–1278. doi:10.1257/aer.102.4.1241

Dupas, P. (2014). Short-Run Subsidies and Long-Run Adoption of New Health Products: Evidence From a Field Experiment. (2014). *Econometrica, 82(1), 197–228.* doi:10.3982/ecta9508

Egami, Hiroyuki & Matsumoto, Tomoya. (2020). Mobile Money Use and Healthcare Utilization: Evidence from Rural Uganda. *Sustainability*. 12. 3741. 10.3390/su12093741.

Fischer, G., D. Karlan, M. McConnell and P. Raffler (2019). "Short-term subsidies and seller type: A health products experiment in Uganda." *Journal of Development Economics* 137: 110-124. https://doi.org/10.1016/j.jdeveco.2018.07.013

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. doi:10.2307/3088292

Finan, F., Olken, B. A., & Pande, R. (2017). The Personnel Economics of the Developing

State. *Handbook of Economic Field Experiments*, 467–514. doi:10.1016/bs.hefe.2016.08.001

Fry, K., Firestone, R., & Chakraborty, N. M. (2014). Measuring equity with nationally representative wealth quintiles. Washington DC: Population Services International. Retrieved on September 9, 2021, from https://www.researchgate.net/publication/267327645_Measuring_Equity_with_Nationally_Representative_Wealth_Quintiles

Fryer, R. G. (2013). Teacher Incentives and Student Achievement: Evidence from New York City Public Schools. *Journal of Labor Economics*, 31(2), 373–407. doi:10.1086/667757

Grépin, K. A., Habyarimana, J., & Jack, W. (2019). Cash on delivery: Results of a randomized experiment to promote maternal health care in Kenya. *Journal of health economics*, *65*, 15-30.

Henriksson, D.K., Peterson, S.S., Waiswa, P. *et al.* Decision-making in district health planning in Uganda: does use of district-specific evidence matter?. *Health Res Policy Sys* **17,** 57 (2019). https://doi.org/10.1186/s12961-019-0458-6.

Hunter BM, Harrison S, Portela A, Bick D (2017) The effects of cash transfers and vouchers on the use and quality of maternity care services: A systematic review. PLoS ONE 12(3): e0173068. doi:10.1371/journal.pone.0173068

Jehan, K., Sidney, K., Smith, H., & De Costa, A. (2012). Improving access to maternity services: an overview of cash transfer and voucher schemes in South Asia. *Reproductive health matters*, *20*(39), 142-154.

Kawungezi, P. C., AkiiBua, D., Aleni, C., Chitayi, M., Niwaha, A., Kazibwe, A., . . . Tukei, C. (2015). Attendance and utilization of antenatal care (ANC) services: multi-center study in upcountry areas of Uganda. *Open journal of preventive medicine*, *5*(3), 132.

Keya, K. T., Bellows, B., Rob, U., & Warren, C. (2018). Improving Access to Delivery Care and Reducing the Equity Gap Through Voucher Program in Bangladesh: Evidence From Difference-in-Differences Analysis. *International quarterly of community health education*, *38*(2), 137-145.

Kremer, M., Rao, G., Schilbach, F. (2019) *Handbook of Behavioral Economics*, vol2, Chapter 5.

Lagarde M, Haines A, Palmer N. (2009). The impact of conditional cash transfers on health outcomes and use of health services in low and middle income countries. Cochrane Database of Systematic Reviews, Issue 4. Art. No.: CD008137. DOI: 10.1002/14651858.CD008137.

Lang'at, E., Mwanri, L. Healthcare service providers' and facility administrators' perspectives of the free maternal healthcare services policy in Malindi District, Kenya: a qualitative study. *Reprod Health* 12, 59 (2015). https://doi.org/10.1186/s12978-015-0048-1

Matsuoka, S., Aiga, H., Rasmey, L. C., Rathavy, T., & Okitsu, A. (2010). Perceived barriers to utilization of maternal health services in rural Cambodia. *Health policy*, *95*(2-3), 255-263.

Miller, G., Luo, R., Zhang, L., Sylvia, S., Shi, Y., Foo, P., … Rozelle, S. (2012). Effectiveness of provider incentives for anaemia reduction in rural China: a cluster randomised trial. *BMJ, 345(jul27 1), e4809–e4809.* doi:10.1136/bmj.e4809

Ministry of Health (2015). Health Sector Development Plan 2015/2016 - 2019/2020

Mohanan M, Bauhoff S, Forgia GL, et al. (2014). Effect of Chiranjeevi Yojana on institutional deliveries and neonatal and maternal outcomes in Gujarat, India: a difference-in-differences analysis. *Bull World Health Organ;* 92: 187–194.

MoH, U. (2009). Road map for accelerating the reduction of maternal and neonatal mortality and morbidity in Uganda 2006Á2015. *Kampala: Uganda Ministry of Health*.

MoH HF Master List (2018). Ministry of Health, National Health Facility Master List 2018. A complete list of all Health Facilities in Uganda. Retrieved from www.moh.go.ug on February 25, 2021.

Nguyen, H. T., Hatt, L., Islam, M., Sloan, N. L., Chowdhury, J., Schmidt, J.-O., . . . Wang, H. (2012). Encouraging maternal health service utilization: an evaluation of the Bangladesh voucher program. *Social science & medicine*, *74*(7), 989-996.

Nsasira, R., Basheka, B. C., & Oluka, P. N. (2013). Public private partnerships (PPPs) and enhanced service delivery in Uganda: implications from the energy sector. *International Journal of Business Administration*, *4*(3), 48.

Nyqvist, M. B., Guariso, A., Svensson, J., & Yanagizawa-Drott, D. (2019). Reducing Child Mortality in the Last Mile: Experimental Evidence on Community Health Promoters in Uganda. *American Economic Journal*: Applied Economics, 11(3), 155–192. https://doi:10.1257/app.20170201

Obare, F., Okwero, P., Villegas, L., Mills, S., & Bellows, B. (2016). Increased coverage of maternal health services among the poor in western Uganda in an output-based aid voucher scheme. *The World Bank.*

Rafa, M., Moyer, J.D., Wang, X., Sutton, P., 2017. Estimating District GDP in Uganda. Kampala. Retrieved from https://pardee.du.edu/sites/default/files/Estimating%20District%20GDP%20in%20Uganda%20.pdf on April 9, 2021.

Riaz, A., Zaidi, S., & Khowaja, A. R. (2015). Perceived barriers to utilizing maternal and neonatal health services in contracted-out versus government-managed health facilities in the rural districts of Pakistan. *International journal of health policy and management, 4*(5), 279.

Skiles M, Curtis S, Basinga P, et al. (2013). An equity analysis of performance-based financing in Rwanda: are services reaching the poorest women?. *Health Pol Plan*; 28: 825–837.

Strachan, D. L., Kirkwood, B., Meek, S. R., Tibenderana, J., Källander, K., Hill, Z., … Conteh, L. (2012). Interventions to Improve Motivation and Retention of Community Health Workers Delivering Integrated Community Case Management (iCCM): Stakeholder Perceptions and Priorities. *The American Journal of Tropical Medicine and Hygiene*, 87(5_Suppl), 111–119. https://doi:10.4269/ajtmh.2012.12-0030

Tama E, Molynuex S, Waweru E, et al. (2017) Examining the implementation of the free maternity services policy in Kenya: A qualitative process evaluation. *International Journal of Health Policy Management*. In Press. 2017:1–11.

Uganda Bureau of Statistics - UBOS and ICF. 2018. Uganda Demographic and Health Survey 2016. Kampala, Uganda and Rockville, Maryland, USA: UBOS and ICF. https://dhsprogram.com/publications/publication-fr333-dhs-final-reports.cfm

UBOS Statistcal Abstract (2019). Uganda Bureau of ststistics 2019 Statistical Abstract. Retrieved from www.ubos.org on February 25, 2021

UBOS Statistcal Abstract (2020). Uganda Bureau of ststistics 2020 Statistical Abstract. Retrieved from www.ubos.org on February 15, 2022

UDHS (2011). Uganda Bureau of Statistics (UBOS) and ICF International Inc. 2012. Uganda Demographic and Health Survey 2011. Kampala, Uganda: UBOS and Calverton, Maryland: ICF International Inc.

UDHS (2016). Uganda Bureau of Statistics (UBOS) and ICF. 2018. *Uganda Demographic and Health Survey 2016*. Kampala, Uganda and Rockville, Maryland, USA: UBOS and ICF

UNCST (2014). Uganda National Council for Science and Technology (UNCST). National Guidelines for Research involving Humans as Research Participants. Kampala, Uganda: UNCST

UN Women (2018). UN Women's global monitoring report. *Turning Promises into Action: Gender Equality in the 2030 Agenda for Sustainable Development*.

Van de Poel, E., Flores, G., Ir, P., & Van Doorslaer, E. (2014). Can vouchers deliver? An evaluation of subsidies for maternal health care in Cambodia. *Bulletin of the World Health Organization, 92*, 331-339.

Wagner, Z., Asiimwe, J. B., & Levine, D. I. (2019). When financial incentives backfire: Evidence from a community health worker experiment in Uganda. *Journal of Development Economics*, 102437. doi:10.1016/j.jdeveco.2019.102437

World Bank (2014). Project Appraisal Document. Uganda Reproductive Health Voucher Project (RHVP-II). Retrieved from file:///C:/GRIPS/Reproductive%20Health/World%20Bank%20Appraisal%20Report.pdf on June 22, 2019.

World Health Organization. (2016). Global strategy on human resources for health: workforce 2030. *World Health Organization*. https://apps.who.int/iris/handle/10665/250368

World Health Organization (WHO) (2018). *World health statistics 2018*: monitoring health for the Sustainable Development Goals.

World Health Organization (WHO) (2019). *World health statistics 2019*: monitoring health for the Sustainable Development Goals.

Yamauchi, C., & Manang, F. E. (2019). The Impact of Access to Health Facilities on Maternal Care Use, Travel Patterns and Health Status: Evidence from Longitudinal Data from Uganda. *Economic Development and Cultural Change*. doi:10.1086/702794

TABLE 2.1: DESCRIPTIVE STATISTICS FOR HEALTH FACILITIES

	All	Non-program facilities	Program facilities	t-stats	p-value
	(1)	(2)	(3)	(4)	(5)
Ownership, level, and proximity of health facility					
1=Public health facility	0.643	0.708	0.529	2.084	0.040
	[0.481]	[0.457]	(0.504]		
1=Higher level facility (HCIII and above)	0.679	0.607	0.804	-2.575	0.011
	[0.469]	[0.491]	[0.401]		
Distance to DHO in km	17.759	18.834	15.884	1.338	0.183
	[14.120]	[15.915]	[10.143]		
Distance to nearest referral facility in km	21.481	19.971	24.118	-1.323	0.189
	[17.878]	[17.812]	[17.860]		
1=Foot common transport mode to referral	0.021	0.011	0.039	-0.943	0.349
	[0.145]	[0.106]	[0.196]		
1=Vehicle common transport mode to referral	0.314	0.292	0.353	-0.731	0.466
	[0.466]	[0.457]	[0.483]		
1=Cycle common transport mode to referral	0.664	0.697	0.608	1.049	0.297
	[0.474]	[0.462]	[0.493]		
Travelling time to referral facility in mins	45.150	43.258	48.451	-0.783	0.436
	[36.314]	[34.420]	[39.541]		
CHWs					
1=CHWs have been attached to facility	0.993	0.989	1.000	-1.000	0.320
	[0.085]	[0.106]	[0.000]		
Number of observations	140	89	51		

Notes: Standard deviations in brackets and t-statistics for testing means between non-program and program facilities.

TABLE 2.2: DESCRIPTIVE STATISTICS FOR USAGE OF MATERNAL CARE SERVICES, NUMBER OF BEDS AND STAFF AND QUALITY OF SERVICES AT HEALTH FACILITY

	Pre-Program				Program Implementation period (post1)				Program withdrawal period (post2)			
Outcome variables	All (1)	Non-Program (2)	Program (3)	p-value (4)	All (5)	Non-Program (6)	Program (7)	p-value (8)	All (9)	Non-Program (10)	Program (11)	p-value (12)
Usage of maternal care services												
Average number of women who took 1st ANC visit	794.282 [665.889]	783.803 [705.231]	814.023 [592.120]	0.801	789.243 [693.309]	665.841 [639.597]	1019.775 [731.697]	0.000	727.167 [670.103]	644.761 [632.070]	872.200 [715.875]	0.065
Average number of women who took 4th ANC visit	281.975 [246.611]	266.663 [226.516]	311.143 [281.586]	0.380	312.066 [274.036]	244.257 [243.237]	438.691[a] [283.872]	0.000	316.435 [300.663]	268.750 [270.246]	400.360 [334.370]	0.020
Average number of women who delivered in facility	443.846 [477.620]	436.103 [436.886]	458.472 [552.821]	0.834	512.303 [540.846]	413.926 [507.311]	666.663[a] [556.932]	0.000	530.465 [669.461]	428.205 [583.749]	693.245 [765.366]	0.041
Average number of women who took PNC	280.837 [341.477]	261.030 [311.010]	316.703 [392.732]	0.460	1004.518 [1054.983]	780.471[a] [859.985]	1365.209[a] [1228.608]	0.000	1374.597 [1576.352]	1215.612[bc] [1450.638]	1634.163[b] [1747.011]	0.163
Number of Beds and staff												
Average number of inpatient medical general beds	11.007 [18.656]	8.851 [17.614]	14.917 [20.010]	0.082	12.800 [20.930]	10.135 [18.459]	17.451 [24.153]	0.065	14.443 [22.639]	11.899 [19.710]	18.882 [26.634]	0.106
Average number of inpatient maternity beds	5.649 [7.819]	4.616 [6.412]	7.500 [9.654]	0.068	6.264 [7.735]	5.079 [7.102]	8.333 [8.409]	0.022	7.300 [8.786]	5.843 [7.198]	9.843 [10.631]	0.019
Average number of delivery beds	1.254 [1.485]	1.103 [1.570]	1.532 [1.283]	0.092	1.414 [1.705]	1.112 [1.526]	1.941 [1.881]	0.009	1.693 [1.403]	1.483 [1.610]	2.059[b] [0.835]	0.006
Average number of medical doctors	0.422 [1.061]	0.345 [1.032]	0.563 [1.109]	0.266	0.550 [1.127]	0.360 [0.944]	0.882 [1.336]	0.016	0.671 [1.322]	0.528 [1.289]	0.922 [1.354]	0.095
Average number of nurses	5.657 [10.738]	5.860 [12.700]	5.292 [5.864]	0.724	6.614 [11.319]	6.640 [13.245]	6.569 [6.886]	0.966	6.800 [11.696]	6.955 [13.775]	6.529 [6.798]	0.807
Average number of midwives	2.507 [3.589]	2.116 [3.609]	3.208 [3.482]	0.089	3.271 [4.044]	2.573 [3.896]	4.490 [4.047]	0.007	3.486 [4.003]	2.921 [3.955]	4.471 [3.931]	0.027
Quality of services												
Average number of times per day inpatient is served	1.496 [2.140]	1.108 [1.431]	2.196 [2.918]	0.021	2.373 [5.556]	1.965 [6.538]	3.082 [3.148]	0.186	1.955 [2.341]	1.500 [1.847]	2.750 [2.870]	0.008
Average waiting time (minutes) for patient to be served	8.305 [9.085]	7.122 [8.656]	10.413 [9.540]	0.056	9.887 [9.696]	8.452 [10.322]	12.347 [8.033]	0.017	10.218 [9.500]	9.177 [10.435]	12.063 [7.309]	0.065
1=Very Good ANC services	0.234 [0.425]	0.193 [0.397]	0.311 [0.468]	0.154	0.415 [0.495]	0.282 [0.453]	0.640[a] [0.485]	0.000	0.396 [0.491]	0.375[b] [0.487]	0.431[c] [0.500]	0.519
1=Very Good Delivery services	0.269 [0.446]	0.259 [0.442]	0.283 [0.455]	0.787	0.448 [0.499]	0.308 [0.465]	0.627[a] [0.488]	0.001	0.437 [0.498]	0.412 [0.496]	0.471 [0.504]	0.527
1=Very Good PNC services	0.202 [0.403]	0.177 [0.384]	0.244 [0.435]	0.391	0.359 [0.481]	0.247 [0.434]	0.540[a] [0.503]	0.001	0.421 [0.496]	0.410[bc] [0.495]	0.440[b] [0.501]	0.735
Number of observations	140	89	51		140	89	51		140	89	51	

Notes: Standard deviations in brackets and p-values for testing means between non-program and program facilities. The outcomes on usage of maternal care services are from MOH administrative data while the others (number of beds and staff, quality of services) are from URHVP Health Facility Survey. Very good service takes 1 for responses "Very Good" and 0 for responses "Very Poor or Poor or Fair or Good". Significant differences over time ($p<0.05$) in the outcome variables of each study group between periods pre and post1 (marked [a]) in columns 6 and 7; between periods pre and post 2 (marked [b]) in columns 10 and 11; and between periods post1 and post2 (marked [c]) in columns 10 and 11. Significance levels at $p<0.05$.

TABLE 2.3: DESCRIPTIVE STATISTICS FOR WOMEN

	All	Non-program women	Program women	t-stats	p-value
	(1)	(2)	(3)	(4)	(5)
Basic information					
Age	31.931	31.579	32.577	-1.390	0.166
	[5.621]	[5.512]	(5.789]		
Religion (1=Christian)	0.829	0.820	0.845	-0.537	0.592
	[0.377]	[0.385]	(0.363]		
Marital status (1=Married/Living with partner)	0.935	0.916	0.969	-1.950	0.052
	[0.248]	[0.279]	(0.174]		
Distance from home to facility (km)	3.029	2.546	3.917	-2.257	0.025
	[4.487]	[3.968]	(5.215]		
Common travel mode to facility (1=walk)	0.724	0.736	0.701	0.610	0.543
	[0.448]	[0.442]	[0.460]		
Travel time from home to facility (mins by common mode of walking)	28.970	25.550	35.559	-1.678	0.097
	[35.048]	[28.370]	(44.738]		
Reproduction					
Number of biological children living with woman	4.040	3.961	4.186	-0.986	0.326
	[1.747]	[1.672]	[1.878]		
Times pregnant in last 10 years	3.087	3.090	3.082	0.054	0.957
	[1.046]	[1.004]	[1.124]		
Methods to avoid/delay pregnancy (1=aware)	0.975	0.989	0.948	1.685	0.095
	[0.158]	[0.106]	[0.222]		
Methods to avoid/delay pregnancy (1=used any)	0.876	0.888	0.856	0.743	0.458
	[0.330]	[0.317]	[0.353]		
Education & Information access					
Attend school (1=woman attend school)	0.931	0.916	0.959	-1.478	0.141
	[0.254]	[0.279]	[0.200]		
Level of educ (1=woman completed at least primary level)	0.527	0.546	0.495	0.789	0.431
	[0.500]	[0.499]	[0.503]		
Attend school (1=spouse attend school)	0.933	0.919	0.958	-1.344	0.180
	[0.250]	[0.273]	[0.201]		
Level of educ (1=spouse completed at least primary level)	0.645	0.642	0.652	-0.170	0.865
	[0.479]	[0.481]	[0.479]		
1=woman can read, write, speak local language	0.487	0.500	0.464	0.570	0.569
	[0.501]	[0.501]	[0.501]		
1=woman has high proficiency in reading English	0.193	0.174	0.227	-1.025	0.307
	[0.395]	[0.380]	[0.421]		
1=woman has high proficiency in writing English	0.156	0.140	0.186	-0.950	0.344
	[0.364]	[0.348]	[0.391]		
1=woman has high proficiency in speaking English	0.156	0.152	0.165	-0.285	0.776
	[0.364]	[0.360]	[0.373]		
1=woman often read newspaper in last 12 months	0.018	0.017	0.021	-0.216	0.829
	[0.134]	[0.129]	[0.143]		
1=woman often listened to radio in last 12 months	0.785	0.781	0.794	-0.250	0.803
	[0.411]	[0.415]	[0.407]		
1=woman often watched TV in last 12 months	0.244	0.287	0.165	2.389	0.018
	[0.430]	[0.453]	[0.373]		
1=woman has used the internet	0.062	0.090	0.010	3.338	0.001
	[0.241]	[0.287]	[0.102]		
Number of observations	275	178	97		

Notes: Standard deviations in brackets and t-statistics for testing means between non-program and program women. High proficiency in reading, writing, and speaking English equals 1 for responses "Excellent or Good" and 0 for responses "Fair or Poor or Not able". Often read a newspaper, listened to radio, watched TV, and used internet in last 12 months equals 1 for responses "Almost every day or at least once a week" and 0 for responses "Less than once a week or Not at all".

TABLE 2.4: DESCRIPTIVE STATISTICS FOR USAGE OF MATERNAL CARE SERVICES, QUALITY OF SERVICES AND MATERNAL AND CHILD OUTCOMES FOR WOMEN

	Pre-Program				Program Implementation period (post)			
Outcome variables	All (1)	Non-Program (2)	Program (3)	p-value (4)	All (5)	Non-Program (6)	Program (7)	p-value (8)
Usage of maternal care services								
1=Woman went to a facility for ANC	0.983 [0.129]	0.987 [0.114]	0.976 [0.153]	0.498	0.995 [0.068]	0.993 [0.085]	1.000 [0.000]	0.158
Average number of ANC visits per pregnancy	4.168 [1.397]	4.248 [1.433]	4.024 [1.323]	0.142	4.220 [1.338]	4.223 [1.330]	4.214 [1.356]	0.943
1=Woman went to facility for 4+ ANC visit per pregnancy	0.741 [0.440]	0.750 [0.440]	0.724 [0.449]	0.596	0.817 [0.387]	0.809 [0.394]	0.830[a] [0.377]	0.577
1=Woman went to facility for delivery	0.788 [0.409]	0.754 [0.431]	0.850 [0.358]	0.024	0.896 [0.306]	0.849[a] [0.358]	0.980[a] [0.140]	0.000
1= Woman went to facility for PNC (6-8 weeks after birth)	0.638 [0.481]	0.628 [0.484]	0.656 [0.477]	0.601	0.701 [0.458]	0.647 [0.479]	0.799[a] [0.402]	0.001
Quality of services								
1=Very Good CHW services	0.244 [0.431]	0.234 [0.426]	0.262 [0.445]	0.739	0.457 [0.499]	0.262 [0.441]	0.664[a] [0.474]	0.000
1=Very Good Health facility services	0.314 [0.465]	0.314 [0.465]	0.314 [0.466]	0.998	0.509 [0.501]	0.408[a] [0.492]	0.691[a] [0.464]	0.000
1=Very Good Medical staff services	0.399 [0.491]	0.379 [0.486]	0.441 [0.499]	0.298	0.566 [0.496]	0.478[a] [0.501]	0.725[a] [0.448]	0.000
Maternal and child outcomes								
1=Complication during delivery	0.083 [0.276]	0.072 [0.259]	0.102 [0.304]	0.342	0.106 [0.309]	0.063 [0.243]	0.185 [0.390]	0.001
1=Complication during delivery and had Assisted and C/section delivery	0.050 [0.217]	0.038 [0.192]	0.071 [0.258]	0.210	0.050 [0.217]	0.018 [0.135]	0.106 [0.309]	0.001
1=Baby weighed at birth	0.788 [0.409]	0.750 [0.434]	0.858 [0.350]	0.010	0.861 [0.347]	0.798 [0.402]	0.974[a] [0.161]	0.000
Average baby birth weight (Kgs)	3.573 [0.877]	3.573 [0.894]	3.572 [0.854]	0.997	3.583 [1.781]	3.592 [2.162]	3.570 [0.985]	0.898
1=Baby weight <2.5kgs	0.073 [0.261]	0.079 [0.271]	0.064 [0.246]	0.633	0.082 [0.275]	0.083 [0.276]	0.082 [0.275]	0.964
1=pregnancy resulted into livebirth	0.965 [0.184]	0.971 [0.168]	0.954 [0.210]	0.432	0.941 [0.235]	0.947 [0.224]	0.931 [0.254]	0.509
1=pregnancy resulted into stillbirth	0.005 [0.073]	0.004 [0.064]	0.008 [0.087]	0.689	0.018 [0.133]	0.011 [0.102]	0.031 [0.175]	0.171
1=pregnancy resulted into miscarriage	0.019 [0.136]	0.017 [0.128]	0.023 [0.150]	0.685	0.027 [0.162]	0.032 [0.175]	0.019 [0.136]	0.388
1=pregnancy resulted into neonatal death	0.011 [0.103]	0.008 [0.091]	0.015 [0.123]	0.570	0.011 [0.106]	0.011 [0.102]	0.013 [0.111]	0.857
Number of observations	389	241	148		427	284	143	

Notes: Standard deviations in brackets and p-values for testing means between non-program and program women. The outcomes are from URHVP Women Survey. Very good service equals 1 for responses "Very Good" and 0 for responses "Very Poor or Poor or Fair or Good". Significant differences over time ($p<0.05$) in the outcome variables of each study group between periods pre and post1 (marked [a]) in columns 6 and 7. Significance levels at $p<0.05$.

TABLE 2.5: IMPACT OF URHVP ON UTILIZATION OF MATERNAL HEALTH CARE SERVICES AT HEALTH FACILITIES

	Log of number of women who take:							
Panel A	(1) ANC1	(2) ANC4	(3) Delivery	(4) PNC	(5) ANC1	(6) ANC4	(7) Delivery	(8) PNC
Program (program_t1)	**0.226**	**0.378***	**0.855****	**0.614***	**0.244**	**0.381****	**0.766****	**0.563***
	(0.200)	**(0.162)**	**(0.227)**	**(0.255)**	**(0.145)**	**(0.137)**	**(0.204)**	**(0.273)**
Post-Program (program_t2)	**0.067**	**0.141**	**0.693****	**0.511**	**0.109**	**0.162**	**0.405**	**0.366**
	(0.231)	**(0.219)**	**(0.245)**	**(0.316)**	**(0.239)**	**(0.239)**	**(0.294)**	**(0.501)**
Constant	6.955**	6.057**	6.412**	5.970**	6.308**	5.510**	5.899**	6.072**
	(0.202)	(0.167)	(0.201)	(0.233)	(0.084)	(0.082)	(0.126)	(0.154)
R-Wolf p-value (program_t1)	[0.297]	[0.109]	[0.020]	[0.010]	[0.010]	[0.010]	[0.010]	[0.010]
R-Wolf p-value (program_t2)	[0.891]	[0.465]	[0.079]	[0.010]	[0.277]	[0.208]	[0.109]	[0.178]
Obs.	784	779	696	708	784	779	696	708
Facility Fixed Effects	Y	Y	Y	Y	Y	Y	Y	Y
Year Fixed Effects	Y	Y	Y	Y	Y	Y	Y	Y
District Time Trends	N	N	N	N	Y	Y	Y	Y
t-test (p-value)								
Program_t1= Program_t2	0.115	0.018	0.093	0.528	0.549	0.228	0.113	0.547
log (control group mean) in pre-program period					6.134	5.116	5.389	4.889
					(1.318)	(1.259)	(1.516)	(1.458)
Number of observations					124	122	104	104

Panel B	(1) 1 if took ANC	(2) 1 if took ANC 4+	(3) 1 if Facility Delivery	(4) 1 if took PNC	(5) 1 if took ANC	(6) 1 if took ANC 4+	(7) 1 if Facility Delivery	(8) 1 if took PNC
Program (program_t)	**0.020**	**-0.027**	**0.108**	**0.141**	**0.018**	**-0.036**	**0.152***	**0.118**
	(0.023)	**(0.081)**	**(0.065)**	**(0.080)**	**(0.018)**	**(0.114)**	**(0.076)**	**(0.107)**
Constant	1.098**	-0.301	0.401	0.260	1.186**	0.474	-2.079	-0.097
	(0.139)	(0.554)	(0.677)	(0.617)	(0.372)	(1.445)	(1.073)	(1.247)
R-Wolf p-value (program_t)	[0.426]	[0.852]	[0.198]	[0.198]	[0.743]	[0.960]	[0.158]	[0.178]
Obs.	585	581	598	591	585	581	598	591
Women Fixed Effects	Y	Y	Y	Y	Y	Y	Y	Y
Year Fixed Effects	Y	Y	Y	Y	Y	Y	Y	Y
Controls	Y	Y	Y	Y	Y	Y	Y	Y
District Time Trends	N	N	N	N	Y	Y	Y	Y
Control group mean in pre-program period					0.983	0.741	0.788	0.638
					(0.129)	(0.439)	(0.409)	(0.481)
Number of observations					353	347	363	359

Notes: Outcomes in Panel A are from MoH administrative data - estimates for average marginal effects for number of women who take ANC, delivery, and PNC services in natural log transformation (log(Y+1). Program=1 for all program facilities under the program in year t. Pre-program year (2014), program implementation program_t1 ∈ (2015 - 2018) and post-program withdrawal program_t2 (2019). Regressions 5-8 control for district linear specific time trends. Family-wise p-values (Romano-Wolf P-values), for the regressors (Program_t1 and Program_t2) are reported in square brackets (Clarke et al 2019). T-test (p-value) for statistical difference between coefficients for program period and post-program withdrawal. Standard errors clustered at the district level with health facility fixed effects and year fixed effects. Standard errors in parentheses. Significance levels at * $p < 0.05$, ** $p < 0.01$.

In Panel B, outcomes from URHVP Women Survey - estimates for average marginal effects for likelihood of women taking ANC, delivery, and PNC services at a health facility. Program=1 for all program facilities under the program in year t ∈ (2015-2019). All regressions control for woman's age at pregnancy, dummies for child sex and 1st born child, dummies for primary source of income for woman and spouse, dummies indicating possession of agricultural land, livestock, poultry, and durables (TV, radio, motorcycle, bicycle, house). Regressions 5-8 control for district linear specific time trends. Family-wise p-values (Romano-Wolf P-values), for the regressor (Program_t) are reported in square brackets (Clarke et al 2019). Standard errors clustered at the district level with women fixed effects and year fixed effects. Standard errors in parentheses. Significance levels at * $p < 0.05$, ** $p < 0.01$.

TABLE 2.6: IMPACT OF URHVP ON LIKELIHOOD OF PROVISION OF QUALITY MATERNAL HEALTH CARE SERVICES

	1 if Very Good:			Daily average:		1 if Very Good:			Daily average:	
Panel A	(1) ANC services	(2) Delivery care service	(3) PNC services	(4) Times inpatient served	(5) Waiting time (mins)	(6) ANC services	(7) Delivery care service	(8) PNC services	(9) Times inpatient served	(10) Waiting time (mins)
Implementation period (Post1)	0.087*	0.067	0.078	0.735	0.881	-0.013	0.205*	0.170**	0.612	0.024
	(0.037)	(0.062)	(0.043)	(0.723)	(0.512)	(0.045)	(0.079)	(0.053)	(0.749)	(0.391)
Program × Post1	**0.261***	**0.305****	**0.227***	**-0.191**	**1.420**	**0.298***	**0.326***	**0.307****	**-0.378**	**0.631**
	(0.110)	**(0.109)**	**(0.094)**	**(0.796)**	**(1.290)**	**(0.121)**	**(0.151)**	**(0.102)**	**(0.857)**	**(1.767)**
Withdrawal period (Post2)	0.184**	0.175*	0.251**	0.251*	1.704*	-0.015	0.457**	0.437**	-0.000	-0.000
	(0.048)	(0.066)	(0.040)	(0.100)	(0.646)	(0.075)	(0.101)	(0.054)	(0.000)	(0.000)
Program × Post2	**-0.048**	**0.041**	**-0.045**	**0.320**	**0.496**	**0.031**	**0.085**	**0.126**	**-0.119**	**-1.203**
	(0.107)	**(0.110)**	**(0.077)**	**(0.186)**	**(1.180)**	**(0.151)**	**(0.203)**	**(0.107)**	**(0.299)**	**(2.217)**
Constant	0.232**	0.253**	0.194**	1.598**	8.375**	0.231**	0.257**	0.198**	1.583**	8.371**
	(0.026)	(0.027)	(0.025)	(0.162)	(0.304)	(0.012)	(0.020)	(0.011)	(0.166)	(0.126)
R-Wolf p-value (prog×post1)	[0.010]	[0.010]	[0.010]	[0.317]	[0.109]	[0.139]	[0.139]	[0.139]	[0.832]	[0.832]
R-Wolf p-value (prog×post2)	[0.208]	[0.099]	[0.099]	[0.208]	[0.149]	[0.931]	[0.852]	[0.812]	[0.931]	[0.852]
R.sq.	0.132	0.156	0.171	0.013	0.062	0.220	0.271	0.308	0.021	0.158
Obs.	402	339	388	395	394	402	339	388	395	394
Facility Fixed Effects	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
District Time Trends	N	N	N	N	N	Y	Y	Y	Y	Y
t-test (p-value)										
Post1=Prog×Post1	0.193	0.137	0.229	0.541	0.742	0.058	0.582	0.364	0.533	0.761
Post2=Prog×Post2	0.104	0.415	0.005	0.783	0.472	0.839	0.231	0.064	0.694	0.591
Prog×Post1= Prog ×Post2	0.003	0.015	0.000	0.507	0.247	0.027	0.070	0.031	0.743	0.071
Control group mean in pre-program period						0.193	0.259	0.177	1.108	7.122
						(0.397)	(0.442)	(0.384)	(1.431)	(8.656)
Number of observations						83	58	79	83	82

Panel B	(1) 1 if CHW services are Very good	(2) 1 if HF services are Very good	(3) 1 if med staff services are Very good	(4) 1 if CHW services are Very good	(5) 1 if HF services are Very good	(6) 1 if med staff services are Very good
Program (program_t)	**0.568****	**0.332****	**0.264****	**0.381**	**0.263**	**0.297***
	(0.213)	**(0.087)**	**(0.087)**	**(0.310)**	**(0.162)**	**(0.117)**
Constant	-2.872**	-0.141	-0.170	-1.070	-2.927*	2.924
	(1.009)	(0.462)	(0.809)	(3.484)	(1.492)	(5.050)
R-Wolf p-value (program_t)	[0.010]	[0.010]	[0.010]	[0.317]	[0.238]	[0.139]
Obs.	292	572	519	292	572	519
Women Fixed Effects	Y	Y	Y	Y	Y	Y
Year Fixed Effects	Y	Y	Y	Y	YY	Y
Controls	Y	Y	Y	Y	Y	Y
District time trends	N	N	N	Y	Y	Y
Control group mean in pre-program period				0.244	0.314	0.399
				(0.431)	(0.465)	(0.491)
Number of observations				119	347	308

Notes: Outcomes in Panel A are from URHVP Health Facility Survey; Outcomes in Panel B from URHVP Women Survey. In Panel A, are estimates for the average marginal effects for quality of services (columns 1-3&6-8) and daily average of medical staff care time (columns 4-5&9-10); Panel B, are estimates for average marginal effects for quality of services provided by CHWs, health facility and staff. In Panel B, all regressions control for woman's age at pregnancy, dummy for 1st born dummy, dummies for primary source of income for woman and spouse, dummies indicating possession of agricultural land, livestock, poultry and durables (TV, radio, motorcycle, bicycle, house). Very good level of service provision taking 1 for responses "Very Good" and 0 for responses "Very Poor or Poor or Fair or Good". Regressions in columns 6-10 in Panel A and columns 4-6 in Panel B control for district linear specific time trends. In Panel A and Panel B, Family-wise p-values (Romano-Wolf P-values), for regressors (Program×Post1 and Program×Post2) and (Program_t) respectively are reported in square brackets (Clarke et al 2019). T-test (p-value) for statistical difference between coefficients for period and the interaction term of period with program. In Panel A, standard errors clustered at the district level with health facility level fixed effects; Panel B, standard errors clustered at the district level with women fixed effects and year fixed effects. Standard errors in parentheses. Significance levels at * $p < 0.05$, ** $p < 0.01$.

TABLE 2.7: IMPACT OF URHVP ON QUALITY OF HEALTH FACILITIES (NUMBER OF BEDS AND STAFF)

	Average number of						Average number of					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	Medical General beds	Maternity beds	Delivery beds	M/Doctors	Nurses	Midwives	Medical General beds	Maternity beds	Delivery beds	M/Doctors	Nurses	Midwives
Implementation period (Post1)	1.382*	0.554**	0.016	0.002	0.901**	0.494**	-1.397	-0.055	-0.256**	-0.236*	0.774**	0.628**
	(0.625)	(0.191)	(0.087)	(0.058)	(0.325)	(0.161)	(1.225)	(0.403)	(0.070)	(0.096)	(0.195)	(0.123)
Program × Post1	**2.131**	**0.629**	**0.404**	**0.374****	**0.744**	**0.963****	**1.238**	**0.829**	**0.477**	**0.557****	**1.068***	**1.053****
	(1.437)	**(0.506)**	**(0.306)**	**(0.120)**	**(0.454)**	**(0.298)**	**(2.713)**	**(0.873)**	**(0.308)**	**(0.196)**	**(0.418)**	**(0.332)**
Withdrawal period (Post2)	3.146*	1.318**	0.387**	0.170	1.215*	0.843**	-2.398	0.128	-0.157*	-0.299	0.970**	1.103**
	(1.160)	(0.365)	(0.096)	(0.100)	(0.462)	(0.149)	(2.601)	(0.930)	(0.075)	(0.151)	(0.302)	(0.129)
Program × Post2	**1.798**	**1.374**	**0.151**	**0.245**	**0.390**	**0.595***	**-0.203**	**1.744**	**0.313***	**0.598**	**1.060**	**0.793****
	(2.135)	**(0.897)**	**(0.141)**	**(0.150)**	**(0.591)**	**(0.221)**	**(5.203)**	**(1.860)**	**(0.149)**	**(0.302)**	**(0.604)**	**(0.259)**
Constant	10.760**	5.536**	1.252**	0.415**	5.512**	2.452**	10.716**	5.547**	1.256**	0.416**	5.522**	2.453**
	(0.464)	(0.150)	(0.056)	(0.039)	(0.180)	(0.080)	(0.123)	(0.075)	(0.043)	(0.011)	(0.052)	(0.044)
RWolf p-value(prog×post1)	[0.099]	[0.129]	[0.129]	[0.020]	[0.020]	[0.010]	[0.812]	[0.812]	[0.812]	[0.139]	[0.228]	[0.069]
RWolf p-value(prog×post2)	[0.030]	[0.020]	[0.030]	[0.030]	[0.030]	[0.010]	[0.931]	[0.703]	[0.743]	[0.347]	[0.703]	[0.485]
R.sq.	0.112	0.111	0.073	0.104	0.134	0.292	0.330	0.243	0.123	0.265	0.378	0.391
Obs.	415	414	414	415	414	414	415	414	414	415	414	414
Facility Fixed Effects	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
District Time Trends	N	N	N	N	N	N	Y	Y	Y	Y	Y	Y
t-test (p-value)												
Post1=Prog×Post1	0.688	0.903	0.262	0.027	0.829	0.255	0.506	0.477	0.035	0.010	0.609	0.310
Post2=Prog×Post2	0.659	0.961	0.279	0.749	0.414	0.450	0.780	0.567	0.045	0.057	0.922	0.431
Prog×Post1= Prog ×Post2	0.764	0.360	0.384	0.150	0.245	0.113	0.596	0.462	0.601	0.768	0.982	0.239
Control group mean in pre-program period							8.851	4.616	1.103	0.345	5.860	2.116
							(17.614)	(6.412)	(1.570)	(1.032)	(12.700)	(3.609)
Number of observations							87	86	87	87	86	86

Notes: Outcomes are average number of beds and staff per health facility and are from URHVP Health Facility Survey. Regressions in columns 7-12 control for district linear specific time trends. Family-wise p-values (Romano-Wolf P-values), for the interaction terms (Program×Post1 and Program×Post2) are reported in square brackets (Clarke et al 2019). T-test (p-value) for statistical difference between coefficients for period and the interaction term of period with program. Standard errors clustered at the district level with health facility level fixed effects. Standard errors in parentheses. Significance levels at * $p < 0.05$, ** $p < 0.01$

TABLE 2.8: IMPACT OF URHVP ON MATERNAL AND CHILD HEALTH OUTCOMES

Panel A: Complications & Birthweight	(1) 1 if complication during delivery	(2) 1 if complication during delivery	(3) 1 if baby was weighed	(4) Baby weight (kg)	(5) I if baby birth weight <2.5kgs	(6) 1 if complication during delivery	(7) 1 if complication during delivery	(8) 1 if baby was weighed	(9) Baby weight (kg)	(10) I if baby birth weight <2.5kgs
Program (program_t)	**0.119**	**0.079**	**0.148**	**-0.150**	**-0.063**	**0.062**	**0.035**	**0.194**	**-0.021**	**-0.023**
	(0.070)	**(0.057)**	**(0.082)**	**(0.489)**	**(0.059)**	**(0.137)**	**(0.091)**	**(0.108)**	**(0.276)**	**(0.077)**
Constant	-0.259	-0.388	0.698	3.298*	0.173	1.168	0.758	-1.234	-5.036	0.609
	(0.308)	(0.293)	(0.551)	(1.368)	(0.355)	(1.145)	(0.435)	(1.081)	(4.377)	(0.804)
RWolf p-value(program_t)	[0.158]	[0.158]	[0.119]	[0.931]	[0.426]	[0.762]	[0.762]	[0.188]	[0.822]	[0.822]
Obs.	598	598	599	514	514	598	598	599	514	514
Mother Fixed Effects	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Year Fixed Effects	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Controls	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
District time trends	N	N	N	N	N	Y	Y	Y	Y	Y
Control group mean in pre-program period						0.083	0.050	0.788	3.573	0.073
						(0.276)	(0.217)	(0.409)	(0.877)	(0.261)
Number of observations						363	363	363	286	286

Panel B: Pregnancy outcomes	(1) 1 if livebirth	(2) 1 if stillbirth	(3) 1 if miscarriage	(4) 1 if neonatal death	(5) 1 if livebirth	(6) 1 if stillbirth	(7) 1 if miscarriage	(8) 1 if neonatal death
Program (program_t)	**0.048**	**0.026**	**-0.060**	**-0.014**	**-0.008**	**0.039**	**-0.038**	**0.007**
	(0.056)	**(0.027)**	**(0.035)**	**(0.019)**	**(0.073)**	**(0.027)**	**(0.058)**	**(0.019)**
Constant	1.273**	0.229	-0.336	-0.129	1.342	0.715	-1.167	0.224
	(0.325)	(0.120)	(0.223)	(0.071)	(1.054)	(0.462)	(0.957)	(0.666)
RWolf p-value(program_t)	[0.347]	[0.287]	[0.109]	[0.564]	[0.980]	[0.337]	[0.426]	[0.990]
Obs.	621	621	621	621	621	621	621	621
Mother Fixed Effects	Y	Y	Y	Y	Y	Y	Y	Y
Year Fixed Effects	Y	Y	Y	Y	Y	Y	Y	Y
Controls	Y	Y	Y	Y	Y	Y	Y	Y
District time trends	N	N	N	N	N	N	N	N
Control group mean in pre-program period					0.965	0.005	0.019	0.011
					(0.184)	(0.073)	(0.136)	(0.103)
Number of observations					372	372	372	372

Notes: In Panel A, Outcomes from URHVP Women Survey for women pregnancies in 2010-2019. Regressions (1,2,6&7) on maternal health outcomes control for woman's age at pregnancy, dummies for child sex and 1st born child, dummies for primary source of income for woman and spouse, dummies indicating possession of agricultural land, livestock, poultry and durables (TV, radio, motorcycle, bicycle, house). Regressions 2&7 estimate the incidence of complications at birth for women who had assisted and c-section delivery. Regressions (3-5&8-10) on child health outcomes control for woman's age at pregnancy, dummies for primary source of income for woman and spouse, dummies indication possession of agricultural land, livestock, poultry and durables (TV, radio, motorcycle, bicycle, house). Family-wise p-values (Romano-Wolf P-values), for the regressor (Program_t) are reported in square brackets (Clarke et al 2019). Standard errors clustered at the district level with women fixed effects and year fixed effects. Standard errors in parentheses. Significance levels at * $p < 0.05$, ** $p < 0.01$

In Panel B, outcomes from URHVP Women Survey for women pregnancies in 2010-2019. Regressions control for woman's age at pregnancy, dummies for primary source of income for woman and spouse, dummies indicating possession of agricultural land, livestock, poultry and durables (TV, radio, motorcycle, bicycle, house). Family-wise p-values (Romano-Wolf P-values), for the regressor (Program_t) are reported in square brackets (Clarke et al 2019). Standard errors clustered at the district level with women fixed effects and year fixed effects. Standard errors in parentheses. Significance levels at * $p < 0.05$, ** $p < 0.01$.

TABLE 2.9: HETEROGENEITY IN THE IMPACT OF URHVP ON USAGE OF MATERNAL HEALTH CARE SERVICES AT HEALTH FACILITIES BY PER CAPITA GDP

	Log of women who take			
Panel A	(1) ANC1	(2) ANC4	(3) Delivery	(4) PNC
Program_t1	0.349	0.385	0.919**	0.472
	(0.194)	(0.199)	(0.292)	(0.382)
poor$_d$ × Program_t1	**-0.253**	**-0.021**	**-0.344**	**0.174**
	(0.214)	**(0.236)**	**(0.340)**	**(0.429)**
Program_t2	0.431	0.323	0.709*	0.500
	(0.320)	(0.304)	(0.337)	(0.652)
poor$_d$ × Program_t2	**-0.799**	**-0.406**	**-0.743**	**-0.361**
	(0.453)	**(0.434)**	**(0.484)**	**(1.107)**
Constant	6.345**	5.500**	5.916**	5.990**
	(0.096)	(0.108)	(0.133)	(0.169)
R-Wolf p-value (poord×program_t1)	[0.149]	[0.020]	[0.020]	[0.010]
R-Wolf p-value (poord×program_t2)	[0.762]	[0.762]	[0.624]	[0.624]
Obs.	784	779	696	708
Facility Fixed Effects	Y	Y	Y	Y
Year Fixed Effects	Y	Y	Y	Y
District Time Trends	Y	Y	Y	Y
t-test (p-value)				
program_t1= poord×program_t1	0.112	0.319	0.033	0.693
program_t2= poord×program_t2	0.086	0.277	0.045	0.588
poord×program_t1= poord×program_t2	0.249	0.352	0.325	0.522

Panel B: Mother fixed effects model	(1) 1 if took ANC at facility	(2) 1 if took ANC 4+	(3) 1 if Delivered at facility	(4) 1 if took PNC at facility
Program_t	0.028	-0.163	0.176	0.157
	(0.027)	(0.117)	(0.102)	(0.166)
poor$_d$ × Program_t	**-0.025**	**0.305**	**-0.061**	**-0.099**
	(0.023)	**(0.188)**	**(0.166)**	**(0.173)**
Constant	1.136**	1.082	-2.208	-0.313
	(0.360)	(1.454)	(1.137)	(1.266)
R-Wolf p-value (poord×program_t)	[0.574]	[0.574]	[0.574]	[0.574]
Obs.	585	581	598	591
Women Fixed Effects	Y	Y	Y	Y
Year Fixed Effects	Y	Y	Y	Y
Controls	Y	Y	Y	Y
District time trends	Y	Y	Y	Y
t-test (p-value)				
program_t= poor$_d$×program_t	0.283	0.081	0.340	0.442

Notes: See notes for Table 2.5. Poor district=1 if per Capita GDP is less than median of $196.50 and 0 if otherwise. Median value taken for 30 districts. Family-wise p-values (Romano-Wolf P-values), for the regressors (poord×Program_t1 and poord×Program_t2) in Panel A and (poord × Program_t) in Panel B are reported in square brackets (Clarke et al 2019). In Panel A standard errors clustered at the district level with health facility fixed effects and year fixed effects, while in Panel B standard errors are clustered at the district level with women fixed effects and year fixed effects. Standard errors in parentheses. Significance levels at * $p < 0.05$, ** $p < 0.01$.

TABLE 2.10: HETEROGENEITY IN THE IMPACT OF URHVP ON THE QUALITY OF MATERNAL CARE SERVICES BY PER CAPITA GDP

	1 if Very Good:			Daily average:	
Panel A	(1) ANC services	(2) Delivery care service	(3) PNC services	(4) Times inpatient is served	(5) Waiting time (mins)
Implementation period (Post1)	0.003	-0.190	0.134*	1.202	-0.214
	(0.068)	(0.157)	(0.065)	(1.379)	(0.636)
$Poor_d$ × Post1	-0.014	0.519**	-0.008	-1.313	0.539
	(0.064)	(0.171)	(0.094)	(1.382)	(0.688)
Program × Post1	0.210	0.319	0.236	-1.022	0.928
	(0.191)	(0.215)	(0.126)	(1.585)	(2.641)
$poor_d$ × Program × Post1	**0.183**	**-0.026**	**0.184**	**1.357**	**-0.555**
	(0.237)	**(0.267)**	**(0.196)**	**(1.604)**	**(3.191)**
Withdrawal period (Post2)	0.004	-0.202	0.358**	-0.000	-0.000
	(0.104)	(0.258)	(0.084)	(0.000)	(0.000)
$Poor_d$ × Post2	0.000	0.773*	0.000	0.000	0.000
	(0.000)	(0.282)	(0.000)	(0.000)	(0.000)
Program × Post2	0.046	0.202	0.043	-0.078	-1.385
	(0.208)	(0.258)	(0.130)	(0.418)	(3.200)
$poor_d$ × Program × Post2	**-0.054**	**-0.344**	**0.240**	**-0.231**	**0.608**
	(0.295)	**(0.346)**	**(0.213)**	**(0.569)**	**(3.734)**
Constant	0.231**	0.259**	0.199**	1.577**	8.372**
	(0.012)	(0.016)	(0.012)	(0.174)	(0.123)
R-Wolf p-value (poor×prog×post1)	[0.059]	[0.059]	[0.050]	[0.772]	[0.772]
R-Wolf p-value (poor×prog×post2)	[0.980]	[0.980]	[0.287]	[0.980]	[0.980]
R.sq.	0.229	0.306	0.311	0.031	0.159
Obs.	402	339	388	395	394
Facility Fixed Effects	Y	Y	Y	Y	Y
District Time Trends	Y	Y	Y	Y	Y
t-test (p-value)					
poor×post1= poor×post2	0.823	0.163	0.930	0.350	0.440
prog×post1= poor×prog×post1	0.947	0.456	0.862	0.460	0.792
prog×post2= poor×prog×post2	0.830	0.342	0.530	0.869	0.768
poor×prog×post1=poor×prog×post2	0.289	0.140	0.749	0.275	0.523

Panel B	(1) 1 if CHW services are Very good	(2) 1 if HF services are Very good	(3) 1 if medical staff services are Very good
Program_t	0.440**	0.414*	0.316*
	(0.164)	(0.168)	(0.136)
$poor_d$ × Program_t	**-0.160**	**-0.358**	**-0.045**
	(0.600)	**(0.322)**	**(0.247)**
Constant	-1.269	-3.617**	2.937
	(3.824)	(1.390)	(5.043)
R-Wolf p-value (poord×program_t)	[0.852]	[0.852]	[0.594]
Obs.	292	572	519
Women Fixed Effects	Y	Y	Y
Year Fixed Effects	Y	Y	Y
Controls	Y	Y	Y
District time trends	Y	Y	Y
t-test (p-value)			
program_t= $poor_d$×program_t	0.286	0.078	0.291

Notes: See notes for Table 2.6. Poor district=1 if per Capita GDP is less than median of $196.50 and 0 if otherwise. Median value taken for 30 districts. In Panel A, standard errors clustered at the district level with health facility fixed effects; in Panel B standard errors are clustered at the district level with women facility fixed effects & year fixed effects. Family-wise p-values (Romano-Wolf P-values), for the regressors (poord×Program_t1 and poord×Program_t2) in Panel A and (poord × Program_t) in Panel B are reported in square brackets (Clarke et al 2019). Standard errors in parentheses. Significance levels at * $p < 0.05$, ** $p < 0.01$.

TABLE 2.11: HETEROGENEITY IN THE IMPACT OF URHVP ON QUALITY OF SERVICES OF HEALTH FACILITIES (NUMBER OF BEDS AND STAFF) BY PER CAPITA GDP

	Average number at health facility					
	(1)	(2)	(3)	(4)	(5)	(6)
	Medical General beds	Maternity beds	Delivery beds	M/Doctors	Nurses	Midwives
Implementation period (Post1)	-2.235**	-0.244	-0.100	-0.432**	0.689**	0.589**
	(0.778)	(0.337)	(0.091)	(0.135)	(0.216)	(0.117)
$Poor_d$ × Post1	0.192	0.960*	-0.154	0.019	0.300	0.222
	(0.858)	(0.367)	(0.137)	(0.076)	(0.297)	(0.254)
Program × Post1	0.525	1.373	0.630	0.404*	1.784*	1.624**
	(4.108)	(1.372)	(0.551)	(0.171)	(0.661)	(0.559)
$poor_d$ × Program × Post1	**2.105**	**-1.384**	**-0.390**	**0.475**	**-1.554**	**-1.255**
	(4.299)	**(1.432)**	**(0.592)**	**(0.362)**	**(0.768)**	**(0.644)**
Withdrawal period (Post2)	-3.911**	0.616	0.017	-0.677*	1.070**	1.226**
	(0.638)	(0.301)	(0.101)	(0.257)	(0.296)	(0.161)
$Poor_d$ × Post2	0.000	0.000	0.000	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Program × Post2	-1.780	2.265	0.494*	0.205	1.194	0.944*
	(7.850)	(2.888)	(0.196)	(0.182)	(0.895)	(0.387)
$poor_d$ × Program × Post2	**4.601**	**-1.497**	**-0.528**	**1.148***	**-0.335**	**-0.397**
	(7.953)	**(2.950)**	**(0.281)**	**(0.545)**	**(1.073)**	**(0.504)**
Constant	10.717**	5.547**	1.256**	0.417**	5.522**	2.453**
	(0.125)	(0.073)	(0.043)	(0.011)	(0.052)	(0.046)
R-Wolf p-v ($poor_d$ ×prog×post1)	[0.168]	[0.515]	[0.951]	[0.020]	[0.168]	[0.030]
R-Wolf p-v ($poor_d$ ×prog×post2)	[0.287]	[0.297]	[0.802]	[0.030]	[0.099]	[0.050]
R.sq.	0.333	0.251	0.130	0.295	0.388	0.411
Obs.	415	414	414	415	414	414
Facility Fixed Effects	Y	Y	Y	Y	Y	Y
District Time Trends	Y	Y	Y	Y	Y	Y
t-test (p-value)						
poor×post1= poor×post2	0.825	0.014	0.268	0.804	0.320	0.389
prog×post1= poor×prog×post1	0.851	0.329	0.371	0.880	0.022	0.019
prog×post2= poor×prog×post2	0.688	0.522	0.027	0.145	0.424	0.121
poor×prog×post1=poor×prog×post2	0.552	0.956	0.817	0.014	0.086	0.061

Notes: See notes for Table 2.7. Poor district=1 if per Capita GDP is less than median of $196.50 and 0 if otherwise. Median value taken for 30 districts. Standard errors clustered at the district level with health facility fixed effects. Family-wise p-values (Romano-Wolf P-values), for the interaction terms (poord×Program×Post1 and poord×Program×Post2) are reported in square brackets (Clarke et al 2019). Standard errors in parentheses. Significance levels at * $p < 0.05$, ** $p < 0.01$.

TABLE 2.12: HETEROGENEITY IN THE IMPACT OF URHVP ON MATERNAL AND CHILD HEALTH OUTCOMES BY PER CAPITA GDP

Panel A: Complications and Birthweights	(1) 1 if complication during delivery	(2) 1 if complication during delivery	(3) 1 if baby was weighed	(4) Baby weight (kg)	(5) I if baby birth weight <2.5kgs
Program_t	-0.081	-0.035	0.272	0.120	-0.072
	(0.089)	(0.070)	(0.155)	(0.268)	(0.122)
poor$_d$ × Program_t	**0.364**	**0.179**	**-0.200**	**-0.309**	**0.108**
	(0.258)	**(0.152)**	**(0.188)**	**(0.442)**	**(0.143)**
Constant	1.940	1.139*	-1.669	-5.680	0.834
	(1.339)	(0.557)	(1.136)	(4.772)	(0.760)
R-Wolf. p-value (poor$_d$ ×program_t)	[0.188]	[0.267]	[0.436]	[0.832]	[0.713]
Obs.	598	598	599	514	514
Women Fixed Effects	Y	Y	Y	Y	Y
Year Fixed Effects	Y	Y	Y	Y	y
Controls	Y	Y	Y	Y	Y
District time trends	Y	Y	Y	Y	Y
t-test (p-value)					
program_t= poor$_d$×program_t	0.146	0.255	0.155	0.478	0.486

Panel B: Pregnancy Outcomes	(1) 1 if livebirth	(2) 1 if still birth	(3) 1 if miscarriage	(4) 1 if neonatal death
Program_t	-0.046	0.036	-0.005	0.015
	(0.050)	(0.034)	(0.036)	(0.019)
poor$_d$ × Program_t	**0.095**	**0.008**	**-0.083**	**-0.021**
	(0.105)	**(0.058)**	**(0.064)**	**(0.034)**
Constant	1.547	0.732	-1.345	0.178
	(1.120)	(0.476)	(0.963)	(0.666)
RWolf p-value (poor×Program_t**)**	[0.713]	[0.713]	[0.465]	[0.713]
Obs.	621	621	621	621
Women Fixed Effects	Y	Y	Y	Y
Year Fixed Effects	Y	Y	Y	Y
Controls	Y	Y	Y	Y
District time trends	Y	Y	Y	Y
t-test (p-value)				
prog×post=poor×prog ×post	0.225	0.735	0.209	0.425

Notes: See notes for See notes for Table 2.8. Poor district=1 if per Capita GDP is less than median of $196.50 and 0 if otherwise. Median value taken for 30 districts. Family-wise p-values (Romano-Wolf P-values), for the regressor (poord × Program_t) are reported in square brackets (Clarke et al 2019). Standard errors clustered at the district level with mother fixed effects and year fixed effects. Standard errors in parentheses. Significance levels at * $p < 0.05$, ** $p < 0.01$

TABLE 3.1: DESCRIPTIVE STATISTICS FOR COMMUNITY HEALTH WORKERS (CHWS)

	All	Non-program facilities	Program facilities	p-value
	(1)	(2)	(3)	(5)
Basic information				
Gender (1=Female)	0.533	0.566	0.474	0.150
	[0.500]	[0.497]	(0.502]	
Age	47.471	46.949	48.412	0.275
	[10.306]	[9.968]	(10.878]	
Religion (1=Christian)	0.915	0.914	0.918	0.927
	[0.279]	[0.281]	(0.277]	
Marital status (1=Married/Living with partner)	0.827	0.783	0.907	0.042
	[0.379]	[0.413]	(0.292]	
Residential area (1=Lived in same residence since birth)	0.548	0.537	0.567	0.637
	[0.499]	[0.500]	(0.498]	
Years lived in current residence (if not lived there since birth)	23.797	22.877	25.571	0.147
	[10.250]	[10.735]	[9.104]	
1=Have children and dependents	0.974	0.977	0.969	0.701
	[0.159]	[0.150]	[0.174]	
Number of biological children/dependents	5.509	5.298	5.894	0.107
	[2.681]	[2.423]	[3.074]	
CHW selection, duration, training, and participation				
1=Selected by community popular vote	0.735	0.720	0.763	0.438
	[0.442]	[0.450]	[0.428]	
Number of years been CHW since 1st selection	12.603	12.354	13.052	0.374
	[6.341]	[6.543]	[5.968]	
Number of years been CHW at current facility	10.790	10.800	10.773	0.972
	[5.873]	[5.757]	[6.108]	
Number of health facilities supported before current one	0.202	0.171	0.258	0.126
	[0.429]	[0.407]	[0.463]	
1=CHW received training on appointment	0.952	0.949	0.959	0.699
	[0.214]	[0.222]	[0.200]	
1=CHW received training in the last 10 years	0.974	0.983	0.959	0.287
	[0.159]	[0.130]	[0.200]	
Education & Access to Information				
Level of educ (1=CHW completed at least primary)	0.901	0.863	0.969	0.001
	[0.300]	[0.345]	[0.174]	
Attend school (1=spouse attend school)	0.915	0.909	0.926	0.642
	[0.279]	[0.288]	[0.264]	
Level of educ (1=spouse completed at least primary)	0.717	0.687	0.770	0.160
	[0.451]	[0.465]	[0.423]	
1=CHW can read, write, speak local language	0.915	0.909	0.928	0.575
	[0.279]	[0.289]	[0.260]	
1=CHW has high proficiency in reading English	0.526	0.520	0.536	0.800
	[0.500]	[0.501]	[0.501]	
1=CHW has high proficiency in writing English	0.500	0.474	0.546	0.257
	[0.501]	[0.501]	[0.500]	
1=CHW has high proficiency in speaking English	0.434	0.429	0.443	0.816
	[0.497]	[0.496]	[0.500]	
1=CHW often read newspaper in last 12 months	0.154	0.109	0.237	0.010
	[0.362]	[0.312]	[0.428]	
1=CHW often listened to radio in last 12 months	0.893	0.869	0.938	0.051
	[0.309]	[0.339]	[0.242]	
1=CHW often watched TV in last 12 months	0.371	0.389	0.340	0.428
	[0.484]	[0.489]	[0.476]	
Total Number of observations	272	175	97	

Notes: Standard deviations in brackets and t-statistics for testing means between non-program and program. CHWs. High proficiency in reading, writing, and speaking English equals 1 for responses "Excellent or Good" and 0 for responses "Fair or Poor or Not able". Often read a newspaper, listened to radio, watched TV, and used internet in last 12 months equals 1 for responses "Almost every day or Atleast once a week" and 0 for responses "Less than once a week or Not at all".

TABLE 3.2: DESCRIPTIVE STATISTICS FOR CHW WORK COVERAGE, TIME USE AND SERVICE PROVISION

	Pre-Program				Program Implementation period (post1)				Program withdrawal period (post2)			
Outcomes (continuous variables)	All (1)	Non-Program (2)	Program (3)	p-value (4)	All (5)	Non-Program (6)	Program (7)	p-value (8)	All (9)	Non-Program (10)	Program (11)	p-value (12)
CHW Work coverage - Average												
Average number of villages	5.661 [14.244]	4.390 [13.810]	7.987 [14.817]	0.082	11.260 [23.302]	4.802 [9.430]	22.711[a] [33.915]	0.000	6.868 [16.516]	5.640 [16.691]	9.082[c] [16.043]	0.096
Average number of households	334.248 [845.094]	271.759 [718.872]	448.675 [1032.819]	0.184	1030.97 [3464.372]	357.192 [767.358]	2225.711[a] [5495.518]	0.001	534.478 [1298.153]	402.720 [997.580]	772.186[c] [1692.543]	0.051
CHW Time Use - Average												
1=Farming for 1st major activity away from CHW work	0.858 [0.350]	0.887 [0.318]	0.805 [0.399]	0.126	0.803 [0.398]	0.814 [0.390]	0.784 [0.414]	0.555	0.820 [0.385]	0.840 [0.368]	0.784 [0.414]	0.264
Days per week on CHW work	3.078 [1.748]	3.057 [1.823]	3.117 [1.614]	0.802	3.560 [1.872]	3.125 [1.740]	4.330[a] [1.858]	0.000	3.156 [1.704]	3.174 [1.803]	3.124[c] [1.516]	0.806
Hours per day on CHW work	4.252 [2.197]	4.426 [2.237]	3.935 [2.100]	0.109	5.212 [2.533]	4.901 [2.376]	5.763[a] [2.715]	0.010	4.792 [2.594]	4.960[b] [2.237]	4.490[c] [3.128]	0.193
Hours per week on CHW work	13.495 [11.130]	14.004 [11.587]	12.565 [10.249]	0.346	19.190 [14.270]	15.733 [12.276]	25.320[a] [15.514]	0.000	15.676 [12.711]	16.391 [12.893]	14.387[c] [12.337]	0.208
Days per week on CHW's 1st major activity	5.206 [1.465]	5.184 [1.447]	5.247 [1.506]	0.767	4.933 [1.670]	5.116 [1.589]	4.604[a] [1.768]	0.020	4.886 [1.668]	4.868 [1.714]	4.918 [1.592]	0.811
Hours per day on CHW's 1st major activity	4.782 [1.939]	4.777 [1.689]	4.792 [2.341]	0.959	4.619 [2.219]	4.797 [2.144]	4.302 [2.327]	0.088	4.828 [2.332]	4.779 [2.315]	4.918 [2.370]	0.642
Hours per week on CHW's 1st major activity	25.161 [11.994]	25.043 [10.974]	25.377 [13.742]	0.855	22.627 [12.546]	24.430 [12.777]	19.396[a] [11.493]	0.001	23.435 [13.352]	23.483 [14.167]	23.351[c] [11.820]	0.935
Incentivized Services provided by CHWs – Average days per month for:												
Community health education	6.665 [6.407]	7.128 [6.638]	5.818 [5.909]	0.136	8.546 [7.716]	6.983 [6.493]	11.320[a] [8.885]	0.000	5.673 [5.507]	5.749[b] [5.676]	5.536[c] [5.216]	0.756
Mobilizing health campaigns	4.275 [4.740]	4.447 [5.095]	3.961 [4.021]	0.440	6.223 [7.041]	5.151 [5.875]	8.124[a] [8.438]	0.003	4.985 [5.743]	4.760 [5.528]	5.392[c] [6.120]	0.400
Referring cases to health facilities	9.376 [9.003]	9.475 [9.016]	9.195 [9.035]	0.827	10.376 [9.178]	9.128 [8.734]	12.588[a] [9.567]	0.004	9.184 [8.864]	8.834 [8.912]	9.814[c] [8.788]	0.382
Data collection and reporting	1.927 [2.701]	2.021 [2.722]	1.753 [2.671]	0.483	2.338 [3.647]	2.017 [3.081]	2.907[a] [4.440]	0.082	2.070 [3.415]	1.937 [3.065]	2.309 [3.975]	0.425
Counselling of women	7.239 [7.937]	7.610 [8.368]	6.558 [7.083]	0.328	9.004 [8.850]	7.023 [7.506]	12.515[a] [9.941]	0.000	6.974 [7.661]	6.931 [7.647]	7.052[c] [7.725]	0.902
Other CHW activities	8.138 [9.250]	7.887 [9.161]	8.597 [9.452]	0.592	8.576 [9.508]	7.965 [9.032]	9.660 [10.256]	0.176	7.651 [8.870]	7.451 [8.833]	8.010 [8.972]	0.621
Non-Incentivized Services provided by CHWs – Average days per month for:												
Treating uncomplicated diseases	4.450 [7.445]	4.681 [7.904]	4.026 [6.549]	0.513	5.717 [9.094]	5.791 [9.004]	5.588 [9.297]	0.862	5.441 [9.333]	5.486 [9.335]	5.361 [9.377]	0.916
Providing HIV/AIDS services	6.248 [7.902]	6.766 [8.400]	5.299 [6.848]	0.165	7.862 [9.017]	7.012 [8.634]	9.371[a] [9.517]	0.045	6.092 [8.162]	6.703 [9.114]	4.990[c] [5.968]	0.063
Number of observations	272	175	97		272	175	97		272	175	97	

Notes: Standard deviations in brackets and p-values for testing means between non-program and program CHWs. Significant differences over time ($p<0.05$) in the outcome variables of each study group between periods pre and post1 (marked [a]) in columns 6 and 7; between periods pre and post 2 (marked [b]) in columns 10 and 11; and between periods post1 and post2 (marked [c]) in columns 10 and 11. Significance levels at $p<0.05$.

TABLE 3.3: DESCRIPTIVE STATISTICS FOR PROSOCIAL BEHAVIOR ACTIVITIES (CHARITY, VOLUNTEER WORK AND POLITICAL PARTY ACTIVITIES)

	Pre-Program				Program Implementation period (post1)				Program withdrawal period (post2)			
Outcomes (dummy variables)	All (1)	Non-Program (2)	Program (3)	p-value (4)	All (5)	Non-Program (6)	Program (7)	p-value (8)	All (9)	Non-Program (10)	Program (11)	p-value (12)
1= CHW was involved in Charity activities	0.569 [0.496]	0.553 [0.499]	0.597 [0.494]	0.530	0.735 [0.442]	0.720[a] [0.450]	0.763[a] [0.428]	0.438	0.691 [0.463]	0.669[b] [0.472]	0.732 [0.445]	0.272
1= CHW was involved in Volunteer activities	0.656 [0.476]	0.695 [0.462]	0.584 [0.496]	0.109	0.857 [0.351]	0.840[a] [0.368]	0.887[a] [0.319]	0.276	0.790 [0.408]	0.771 [0.421]	0.825[b] [0.382]	0.289
1= CHW was involved in Political party activities	0.248 [0.433]	0.291 [0.456]	0.169 [0.377]	0.036	0.342 [0.475]	0.354 [0.480]	0.320[a] [0.469]	0.563	0.154 [0.362]	0.143[bc] [0.351]	0.175[c] [0.382]	0.492
Number of observations	272	175	97		272	175	97		272	175	97	

Notes: Standard deviations in brackets and p-values for testing means between non-program and program CHWs. 1 if CHW was involved in an activity at least more than once: Charity activities (blood donation, fundraising, movement for social benefits and other charity work); Volunteer activities (agriculture extension work, awareness & advocacy programs, tree planting and other volunteer activities) and Political party activities (party member campaigns, party agent & vote counting, party mobilization campaigns, party communicator and other party activities). Significant differences over time ($p<0.05$) in the outcome variables of each study group between periods pre and post1 (marked [a]) in columns 6 and 7; between periods pre and post 2 (marked [b]) in columns 10 and 11; and between periods post1 and post2 (marked [c]) in columns 10 and 11. Significance levels at $p<0.05$.

TABLE 3.4: IMPACT OF URHVP ON CHW EFFORTS (COVERAGE THAT CHW USUALLY WORKS)

	Log of number of			
	(1) Villages covered	(2) Households covered	(3) Villages covered	(4) Households covered
During URHVP (post1)	0.177*	0.368**	0.391**	0.503**
	(0.065)	(0.079)	(0.082)	(0.177)
Program × Post1	**0.841****	**1.111****	**0.855****	**1.003****
	(0.147)	**(0.179)**	**(0.199)**	**(0.277)**
Withdrawal URHVP (post2)	0.243**	0.277*	0.655**	0.552
	(0.064)	(0.121)	(0.128)	(0.306)
Program × Post2	**-0.061**	**0.008**	**-0.084**	**-0.162**
	(0.118)	**(0.160)**	**(0.251)**	**(0.415)**
Constant	0.977**	4.437**	0.909*	4.682**
	(0.300)	(0.674)	(0.330)	(0.760)
R-Wolf p-value (prog×post1)	[0.010]	[0.010]	[0.010]	[0.010]
R-Wolf p-value (prog×post2)	[0.139]	[0.139]	[0.960]	[0.960]
R.sq.	0.344	0.345	0.387	0.379
Obs.	593	591	593	591
CHW Fixed Effects	Y	Y	Y	Y
Controls	Y	Y	Y	Y
District time trends	N	N	Y	Y
t-test (p-value)				
post1=program×post1	0.001	0.002	0.073	0.238
post2=program×post2	0.067	0.294	0.043	0.313
program×post1=program×post2	0.000	0.000	0.000	0.000
log (control group mean) in pre-program period			1.125	4.805
			(0 .773)	(1. 220)
Number of observations			141	141

Notes: All outcomes are from the URHVP CHW survey. I estimate the average marginal effects for number of villages and households covered by CHW in natural log transformation (log(Y+1) for full sample. All regressions control for dummies for CHW's primary source of income, possession of agricultural land, livestock, poultry, durables (TV, radio, car, motorcycle, bicycle) and access to financial services (distance to nearest mobile money agent and nearest financial institution). Regressions in columns 3-4 control for district linear specific time trends. Family-wise p-values (Romano-Wolf P-values), for the interaction terms (Program×Post1 and Program×Post2 are reported in square brackets (Clarke et al 2019). T-test (p-value) for statistical difference between coefficients for period and the interaction term for period with program. Standard errors clustered at the district level with CHW fixed effects. Standard errors in parentheses. Significance levels at * $p < 0.05$, ** $p < 0.01$.

TABLE 3.5: IMPACT OF URHVP ON CHW EFFORTS (DAYS PER WEEK, HOURS PER DAY AND HOURS PER WEEK SPENT ON CHW WORK AND CHW'S 1ST MAJOR ACTIVITY)

	Log of number of											
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	Days/week CHW work	Hrs/day CHW work	Hrs/week CHW work	Days/week Major activity	Hrs/day Major activity	Hrs/week Major activity	Days/week CHW work	Hrs/day CHW work	Hrs/week CHW work	Days/week Major activity	Hrs/day Major activity	Hrs/week Major activity
During URHVP	-0.038	0.065*	0.033	0.005	0.034	0.017	-0.060	0.034	-0.030	-0.011	0.005	0.002
	(0.030)	(0.028)	(0.057)	(0.022)	(0.031)	(0.038)	(0.034)	(0.034)	(0.072)	(0.026)	(0.035)	(0.052)
Program × Post1	**0.339****	**0.195****	**0.648****	**-0.124***	**-0.207****	**-0.368****	**0.457****	**0.231****	**0.841****	**-0.071**	**-0.202***	**-0.323***
	(0.073)	**(0.056)**	**(0.122)**	**(0.058)**	**(0.063)**	**(0.107)**	**(0.101)**	**(0.080)**	**(0.183)**	**(0.057)**	**(0.088)**	**(0.127)**
Withdrawal URHVP	-0.009	0.104**	0.108	-0.023	0.050	-0.014	-0.023	0.035	0.010	-0.059	-0.009	-0.047
	(0.033)	(0.035)	(0.056)	(0.025)	(0.031)	(0.039)	(0.045)	(0.057)	(0.096)	(0.052)	(0.052)	(0.079)
Program × Post2	**0.034**	**-0.089**	**-0.062**	**0.023**	**-0.072**	**-0.060**	**0.205**	**-0.009**	**0.252**	**0.129**	**-0.057**	**0.032**
	(0.059)	**(0.056)**	**(0.117)**	**(0.059)**	**(0.053)**	**(0.071)**	**(0.112)**	**(0.107)**	**(0.228)**	**(0.079)**	**(0.093)**	**(0.127)**
Constant	1.419**	1.537**	2.488**	1.743**	1.815**	3.194**	1.405**	1.615**	2.566**	1.756**	1.779**	3.098**
	(0.164)	(0.216)	(0.330)	(0.230)	(0.147)	(0.242)	(0.193)	(0.193)	(0.303)	(0.194)	(0.181)	(0.219)
R-Wolf p-value (prog×post1)	[0.010]	[0.010]	[0.010]	[0.079]	[0.020]	[0.010]	[0.010]	[0.020]	[0.010]	[0.446]	[0.079]	[0.079]
R-Wolf p-value (prog×post2)	[0.960]	[1.000]	[0.980]	[1.000]	[0.960]	[0.822]	[0.436]	[0.980]	[0.545]	[0.436]	[0.842]	[0.901]
R.sq.	0.231	0.193	0.319	0.095	0.104	0.153	0.272	0.241	0.361	0.180	0.147	0.192
Obs.	593	593	593	591	591	591	593	593	593	591	591	591
CHW Fixed Effects	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Controls	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
District time trends	N	N	N	N	N	N	Y	Y	Y	Y	Y	Y
t-test (p-value)												
post1=prog×post1	0.000	0.084	0.000	0.060	0.008	0.004	0.000	0.048	0.001	0.433	0.067	0.049
post2=prog×post2	0.606	0.029	0.294	0.492	0.135	0.651	0.131	0.768	0.430	0.122	0.710	0.678
prog×post1=prog×post2	0.000	0.000	0.000	0.000	0.006	0.000	0.000	0.000	0.000	0.001	0.009	0.000
log (control group mean) in pre-program period							1.298	1.606	2.403	1.782	1.709	3.141
							(0.461)	(0.420)	(0.824)	(0.319)	(0.316)	(0.563)
Number of observations							141	141	141	141	141	141

Notes: All outcomes are from the URHVP CHW survey. I estimate the average marginal effects for number days per week and hours per day spent on CHW work and CHW's major activity in natural log transformation (log(Y+1) for full sample. All regressions control for dummies for CHW's primary source of income, possession of agricultural land, livestock, poultry, durables (TV, radio, car, motorcycle, bicycle) and access to financial services (distance to nearest mobile money agent and nearest financial institution). Regressions in Cols 7-12 control for district linear specific time trends. Family-wise p-values (Romano-Wolf P-values), for the interaction terms (Program×Post1 and Program×Post2) are reported in square brackets (Clarke et al 2019). T-test (p-value) for statistical difference between coefficients for period and the interaction term for period with program. Standard errors clustered at the district level with CHW fixed effects. Standard errors in parentheses. Significance levels at * $p < 0.05$, ** $p < 0.01$.

TABLE 3.6: IMPACT OF URHVP ON CHW EFFORTS (DAYS PER MONTH SPENT ON PROVISION OF HEALTHCARE SERVICES)

	Log number of days per month for provision of:															
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
	Health Education	Health campaigns	Case referrals	Data Collection& Reporting	Women Counselling	Other CHW services	Treatment	HIVAIDS services	Health Education	Health campaigns	Case referrals	Data Collection& Reporting	Women Counselling	Other CHW services	Treatment	HIVAIDS services
During URHVP (post1)	-0.052	0.121**	-0.005	-0.057	0.038	-0.019	-0.096	0.100	-0.097	-0.058	-0.152	-0.471**	-0.085	-0.178*	-0.159	0.129
	(0.065)	(0.038)	(0.099)	(0.053)	(0.115)	(0.054)	(0.184)	(0.051)	(0.101)	(0.074)	(0.089)	(0.061)	(0.121)	(0.070)	(0.125)	(0.102)
Program × Post1	**0.503****	**0.158**	**0.274**	**0.298***	**0.353**	**0.028**	**0.027**	**0.371**	**0.373***	**0.153**	**0.288**	**0.310**	**0.440**	**0.043**	**0.085**	**0.137**
	(0.130)	**(0.111)**	**(0.154)**	**(0.114)**	**(0.173)**	**(0.086)**	**(0.244)**	**(0.206)**	**(0.179)**	**(0.148)**	**(0.171)**	**(0.191)**	**(0.218)**	**(0.121)**	**(0.249)**	**(0.270)**
Withdrawal URHVP (post2)	-0.288**	0.037	-0.107	-0.146	-0.003	-0.066	-0.179	0.054	-0.380*	-0.341**	-0.405**	-0.956**	-0.208	-0.369**	-0.246	0.060
	(0.079)	(0.061)	(0.112)	(0.072)	(0.111)	(0.057)	(0.211)	(0.084)	(0.162)	(0.119)	(0.118)	(0.119)	(0.174)	(0.109)	(0.159)	(0.189)
Program × Post2	**0.049**	**-0.088**	**0.083**	**0.212***	**-0.172**	**-0.013**	**0.036**	**-0.083**	**-0.176**	**-0.054**	**0.202**	**0.249**	**-0.085**	**-0.010**	**0.175**	**-0.388**
	(0.152)	**(0.132)**	**(0.156)**	**(0.098)**	**(0.166)**	**(0.101)**	**(0.265)**	**(0.174)**	**(0.267)**	**(0.222)**	**(0.175)**	**(0.219)**	**(0.248)**	**(0.207)**	**(0.237)**	**(0.317)**
Constant	2.044**	2.006**	1.284*	0.498	2.102**	1.547**	1.579**	1.728**	1.980**	2.120**	1.237*	0.576	1.974**	1.515**	1.067	1.802**
	(0.325)	(0.351)	(0.524)	(0.390)	(0.391)	(0.286)	(0.526)	(0.393)	(0.322)	(0.414)	(0.594)	(0.439)	(0.516)	(0.327)	(0.542)	(0.444)
R-Wolf p-v(prog×post1)	[0.010]	[0.099]	[0.198]	[0.119]	[0.099]	[0.931]	[0.931]	[0.089]	[0.040]	[0.337]	[0.396]	[0.396]	[0.228]	[0.960]	[0.960]	[0.455]
R-Wolf p-v(prog×post2)	[0.713]	[1.000]	[1.000]	[0.980]	[0.921]	[1.000]	[1.000]	[1.000]	[1.000]	[1.000]	[1.000]	[1.000]	[1.000]	[1.000]	[1.000]	[0.861]
R.sq.	0.304	0.165	0.160	0.072	0.151	0.037	0.085	0.118	0.395	0.252	0.244	0.173	0.263	0.127	0.179	0.219
Obs.	593	593	593	593	593	593	593	593	593	593	593	593	593	593	593	593
CHW Fixed Effects	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Controls	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
District time trends	N	N	N	N	N	N	N	N	Y	Y	Y	Y	Y	Y	Y	Y
t-test (p-value)																
Post1=Program×Post1	0.003	0.782	0.228	0.019	0.225	0.703	0.758	0.240	0.093	0.315	0.068	0.002	0.099	0.206	0.488	0.982
Post2=Program×Post2	0.098	0.473	0.441	0.025	0.449	0.690	0.632	0.507	0.628	0.370	0.028	0.001	0.744	0.239	0.244	0.344
Prog×Post1= Prog×Post2	0.000	0.021	0.076	0.383	0.001	0.551	0.938	0.001	0.000	0.105	0.351	0.589	0.003	0.664	0.439	0.003
log (control group mean) pre-program period									1.791	1.393	1.903	0.837	1.675	1.578	0.958	1.463
									(0.783)	(0.752)	(1.045)	(0.690)	(1.033)	(1.173)	(1.190)	(1.129)
Number of observations									141	141	141	141	141	141	141	141

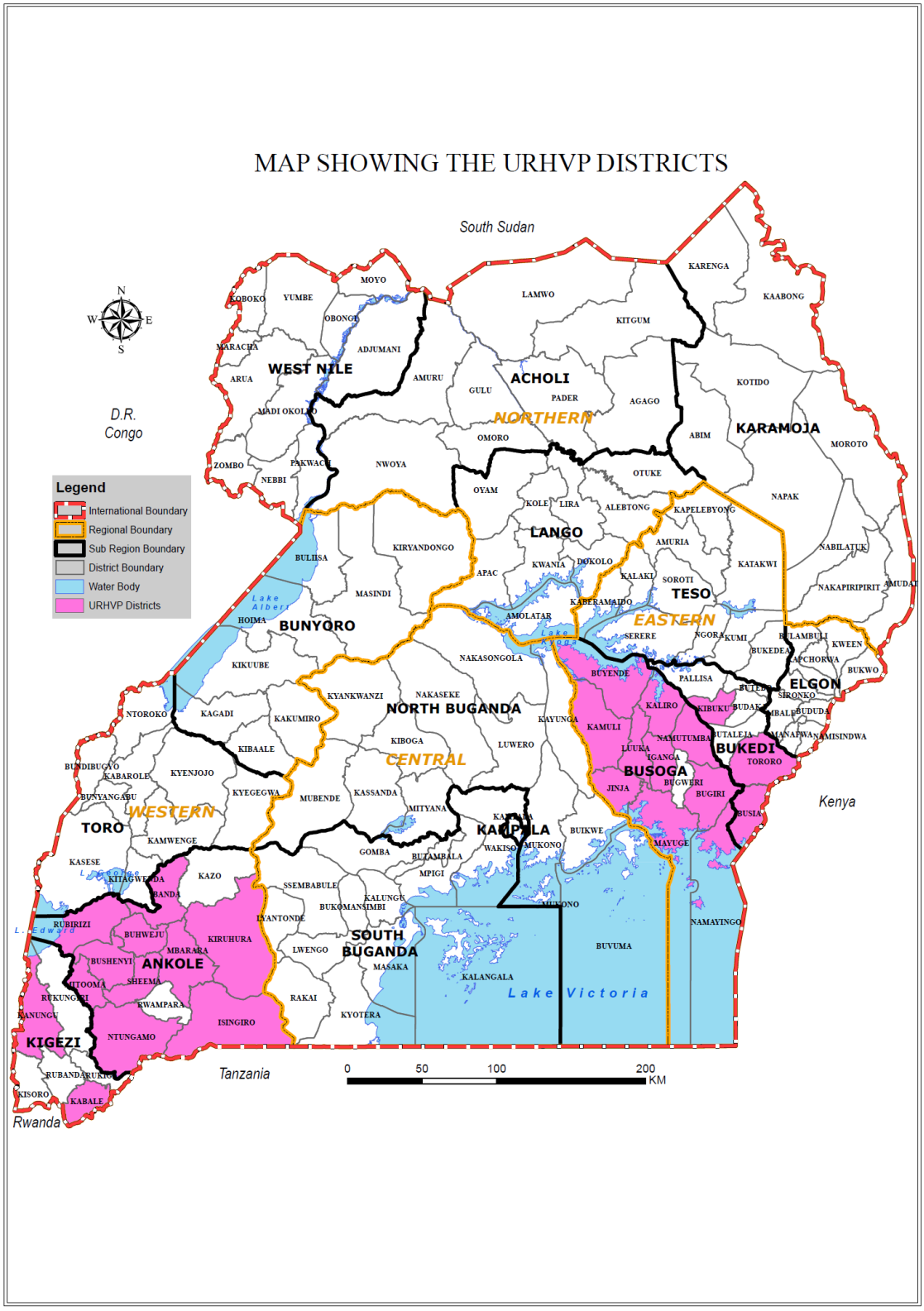
Notes: All outcomes are from the URHVP CHW survey. I estimate the average marginal effects for number days a month CHW spends on provision of various health care services in natural log transformation (log(Y+1) for full sample. Outcomes in Cols 9-16 control for district linear specific time trends. CHW Incentivized services are presented in cols 1-6 and 9-14 while CHW unincentivized services are presented in cols 7-8 and 15-16. All regressions control for dummies for primary source of income for CHW, dummies indicating possession of agricultural land, livestock, poultry, durables (TV, radio, car, motorcycle, bicycle) and access to financial services (distance to nearest mobile money agent and nearest financial institution). Family-wise p-values (Romano-Wolf P-values), for the interaction terms (Program×Post1 and Program×Post2) respectively are reported in square brackets (Clarke et al 2019). T-test (p-value) for statistical difference between coefficients for period and the interaction term for period with program. Standard errors clustered at the district level with CHW fixed effects. Standard errors in parentheses. Significance levels at * $p < 0.05$, ** $p < 0.01$.

TABLE 3.7: IMPACT OF URHVP ON CHW'S INVOLVEMENT IN PROSOCIAL BEHAVIOR ACTIVITIES (CHARITY, VOLUNTEER AND POLITICAL PARTY)

	(1) Charity	(2) Volunteer	(3) Political party	(4) Charity	(5) Volunteer	(6) Political party
During URHVP (post1)	0.107*	0.167**	0.058	0.238*	0.361**	0.209*
	(0.049)	(0.044)	(0.040)	(0.100)	(0.044)	(0.086)
Program × Post1	**-0.009**	**0.032**	**0.086**	**0.029**	**0.137**	**0.087**
	(0.100)	**(0.088)**	**(0.072)**	**(0.184)**	**(0.082)**	**(0.190)**
Withdrawal URHVP (post2)	0.048	0.106**	-0.121*	0.321	0.502**	0.175
	(0.044)	(0.036)	(0.055)	(0.161)	(0.040)	(0.137)
Program × Post2	**0.031**	**0.044**	**0.131**	**0.073**	**0.216***	**0.066**
	(0.095)	**(0.075)**	**(0.092)**	**(0.249)**	**(0.085)**	**(0.318)**
Constant	0.915*	0.655**	0.220	0.958*	0.635**	0.150
	(0.384)	(0.176)	(0.271)	(0.455)	(0.147)	(0.264)
R-Wolf p-value (prog×post1)	[0.257]	[0.040]	[0.129]	[0.376]	[0.030]	[0.178]
R-Wolf p-value (prog×post1)	[0.505]	[0.188]	[0.633]	[0.634]	[0.158]	[0.634]
R.sq.	0.063	0.096	0.153	0.130	0.209	0.227
Obs.	595	595	595	595	595	595
CHW Fixed Effects	Y	Y	Y	Y	Y	Y
Controls	Y	Y	Y	Y	Y	Y
District time trends	N	N	N	Y	Y	Y
t-test (p-value)						
post1=prog×post1	0.394	0.263	0.792	0.430	0.057	0.650
post2=prog×post2	0.893	0.505	0.076	0.519	0.013	0.809
prog×post1=prog×post2	0.466	0.776	0.522	0.629	0.159	0.891
Control group mean (pre-program period)				0.5532	0.6950	0.2908
				(0.4989)	(0.4620)	(0.4557)
Number of observations				141	141	141

Notes: All outcomes are from the URHVP CHW survey. I estimate the average marginal effects for CHW's likelihood of involvement in prosocial behavior activities (Charity, Volunteer and Political party) for full sample. I defined 3 dummies – Charity variable takes the value 1 if CHW was involved in blood donation, fundraising, movement for social benefits and any other charity activities; Volunteer variable takes the value 1 if CHW was involved in agricultural extension work, awareness and advocacy programs, tree planting and any other volunteer activities; and Political party variable takes the value 1 if CHW was involved in political party member campaigns, political party agent and vote counting, political party mobilization campaigns, political party communicator and other political party activities. All regressions control for dummies for primary source of income for CHW, possession of agricultural land, livestock, poultry, durables (TV, radio, car, motorcycle, bicycle) and access to financial services (distance to nearest mobile money agent and nearest financial institution). Regressions in columns 4-6 control for district linear specific time trends. Family-wise p-values (Romano-Wolf P-values), for the interaction terms (Program×Post1 and Program×Post2) are reported in square brackets (Clarke et al 2019). T-test (p-value) for statistical difference between coefficients for period and the interaction term for period with program. Standard errors clustered at the district level with CHW fixed effects. Standard errors in parentheses. Significance levels at * $p < 0.05$, ** $p < 0.01$.

FIGURE 2.1: MAP SHOWING THE URHVP DISTRICTS



Source: Shape files from Uganda Bureau Of Statistics (UBOS)

FIGURE 2.2: SAMPLE OF THE VOUCHER

Source: Marie Stopes Uganda

APPENDIX 2A

TABLE 2A1: COSTS FOR MATERNAL CARE SERVICES

Panel A: URHVP Standardized VSP costs	Materna1 service	Non-public facilities (UGX)	Non-public facilities (USD)	Public facilities (UGX)	Public facilities (USD)
	(1)	(2)	(3)	(4)	(5)
Category A	ANC1	35,000	9.52	26,500	7.14
	ANC2	20,000	5.44	15,000	4.08
	ANC3	19,000	5.17	14,250	3.88
	ANC4	19,000	5.17	14,250	3.88
	PNC	15,000	4.08	11,250	3.06
Category B	Normal delivery	70,000	19.04	52,500	14.28
	Assisted delivery	90,000	24.48	67,500	18.36
	Caesarean delivery	400,000	108.80	300,000	81.60
Category C	Management of simple malaria	23,600	6.42	17,700	4.81
	Management of severe malaria	80,000	21.76	60,000	16.32
	Management of UTI	20,000	5.44	15,000	4.08
Category D	Transport (Emergency referral transport of Ugx2000 @km =USD0.54@km)				
Total with Normal delivery		301,600	82.04	226,200	61.53
Total with Assisted delivery		321,600	87.48	241,200	65.61
Total with Caesarean delivery		631,600	171.80	473,700	128.85
Panel B: Average costs (out-of-pocket expenses)	**Materna1 service**	**Non-public facilities (UGX)**	**Non-public facilities (USD)**	**Public facilities (UGX)**	**Public facilities (USD)**
	(1)	**(2)**	**(3)**	**(4)**	**(5)**
	ANC	9,742.86	2.65	7,375.00	2.01
	Normal delivery	66,446.43	18.07	30,434.25	8.29
	Assisted delivery	71,666.67	19.49	28,333.33	7.71
	Caesarean delivery	469,900.00	127.82	172,727.30	46.98
	Medicines	50,379.31	13.70	23,958.67	6.52
	PNC	3,500.00	0.95	0.00	0.00
	Others (e.g., laboratory tests)	6,684.21	1.82	19,285.71	5.25

Notes: In Panel A, source for the rates in UGX is Marie Stopes International -Uganda. Conversion to USD using exchange rate for September 01, 2015: USD1 = UGX3,676.36, retrieved from https://www.bou.or.ug/bou/collateral/interbank_forms/2015/Sep/Major_01Sep2015.html. Category A: All service costs inclusive of Investigation tests (HIV Test, RPR/TPHA, HB, Urine, Blood Group), Medicines and Consultation (Professional Fees, Counselling & Education, Medical Acumen/ Recording, utilities, space, time). Category B: All service costs inclusive of Consultation (Mama Kit, Professional Fees, Medical Acumen/ Recording, utilities, space, time), Drugs and Post Delivery Care (24hr nursing care, Counselling & Education). Delivery by C-Section only at HCIVs and Hospitals. Category C: All service costs inclusive of Malaria tests and drugs. Total amounts at bottom include category A, respective line item in Category B, and Category C. Non-public facilities refer to private and mission/religious facilities while public facilities refer to government facilities.

In Panel B, source for the rates in UGX is URHVP Women survey, women's pregnancies for 2010-2019. These may be used with caution as they are not representative of all health facilities in Uganda. Conversion to USD using exchange rate for September 01, 2015: USD1 = UGX3,676.36, retrieved from https://www.bou.or.ug/bou/collateral/interbank_forms/2015/Sep/Major_01Sep2015.html. Non-public facilities refer to private and mission/religious facilities while public facilities refer to government facilities. Other out of pocket expenses have been based on where woman went for delivery.

TABLE 2A2: CONSTRUCTION OF SAMPLES AND DATA DESCRIPTION

Panel A: Health facilities sample		
Number of sampled districts	30	
Non-Program districts	12	
Program districts	18	
Number of health facilities in 12 non-program districts	515	
Sampled non-program facilities (a)		66
Number of health facilities in 18 program districts	749	
Program facilities	110	
Sampled program facilities (b)		51
Non-Program facilities	639	
Sampled non-program facilities (c)		23
Total number of sampled health facilities (b+c+a)		140
Total number of observations for 3 periods		420
Panel B: Women sample		
Number of women	275	
Non-Program women from Non-Program facilities		178
Program women from Program facilities		97
Total number of observations for 3 periods	825	
Women's pregnancies for the 10-year period 2010-2019	816	
Pregnancies for non-program women (pre-program)		241
Non-Program women (post-program)		284
Program women (pre-program)		148
Program women (post-program) supported		117
Program women (post-program) unsupported		26
Panel C: Data description	Outcome variables	Data source
Measures for usage of maternal care services	Number of women taking ANC, Delivery and PNC at facility	MoH Annual data 2014-2019
	Likelihood of mother taking ANC, Delivery and PNC at facility during each pregnancy	Women Survey Pregnancy-level data 2010-2019
Measures for quality of service provision	Objective measures - Number of beds (inpatient general, maternity and delivery) and staff (doctors, nurses and midwives), times inpatient is served per day and patient waiting time in minutes. Subjective measure – Provider's self-assessment /positive views on ANC, Delivery and PNC services	Health facility survey data (pre-program period (2011-2014); program period (2015-2019); after withdrawal period (6 months 2019-2020))
	Subjective measure– User's assessment of providers/positive views on CHW, Facility and Medical staff services during each pregnancy	Women Survey Pregnancy-level data 2010-2019
Measures for maternal and child health outcomes	Likelihood of complication at birth, baby birth weight and likelihood of having a livebirth, stillbirth, miscarriage and neonatal death during each pregnancy	Women Survey Pregnancy-level data 2010-2019

Notes: The 30 sampled districts are from 3 regions of Uganda (western, eastern and central). The 12 non-program districts are from western region (3), eastern region (3) and central region (6) while the 18 program districts are from western region (9) and eastern region (9). Out of 110 program facilities in the program districts, one out of every two program health facilities (51 program facilities) were selected and out of 1154 non-program facilities in program and non-program districts, one out of every thirteen non-program facilities were selected (66 non-program facilities from non-program districts and 23 non-program facilities from program districts). URHVP Women Survey targeted women (mothers) with more than 2 pregnancies (at least one in 2011-2014 and at least one in 2015-2019) who took maternal health care at the sampled facility. Survey targeted 2 women from each health facility (2*140=280) however, we had 3 less from program facilities (Kibuku HCIV, Taoky Medical Clinic HCII, Shuuku HCIV), 3 less from non-program facilities (Burambira HCII, Rushaka HCII, Bugobero HCIV) and 1 more from a non-program facility (Budadiri HCIV)

TABLE 2A3: DESCRIPTIVE STATISTICS FOR HEALTH FACILITIES BY WAITING FACILITIES, NUMBER OF STAFF AND CHWS AT FACILITY

	Pre-Program				Program Implementation period (post1)				Program withdrawal period (post2)			
Outcomes (continuous variables)	All (1)	Non-Program (2)	Program (3)	p-value (4)	All (5)	Non-Program (6)	Program (7)	p-value (8)	All (9)	Non-Program (10)	Program (11)	p-value (12)
Waiting facilities at health facility												
Average number of waiting rooms/areas	1.874 [1.933]	1.851 [2.170]	1.917 [1.427]	0.832	2.014 [1.862]	1.966 [2.140]	2.098 [1.253]	0.647	2.293 [1.929]	2.258 [2.198]	2.353 [1.354]	0.754
Average number of seats/beds in waiting rooms/areas	52.700 [50.963]	48.738 [51.884]	60.625 [48.738]	0.221	57.552 [56.963]	54.190 [56.011]	63.200 [58.663]	0.384	64.659 [56.037]	62.188 [56.279]	68.860 [55.938]	0.506
Staff at facility												
Average number of theater assistants	0.207 [0.520]	0.149 [0.471]	0.313 [0.589]	0.103	0.221 [0.524]	0.124 [0.422]	0.392 [0.635]	0.009	0.357 [0.882]	0.292 [0.968]	0.471 [0.703]	0.212
Average number of clinical officers	1.215 [1.181]	1.057 [1.093]	1.500 [1.288]	0.047	1.457 [1.305]	1.236 [1.178]	1.843 [1.433]	0.012	1.543 [1.396]	1.393 [1.320]	1.804 [1.497]	0.107
Average number of anesthetists	0.148 [0.481]	0.115 [0.468]	0.208 [0.504]	0.293	0.236 [0.595]	0.146 [0.490]	0.392 [0.723]	0.034	0.229 [0.580]	0.180 [0.555]	0.314 [0.616]	0.203
CHWs at facility												
Average number of CHWs attached to facility	26.080 [45.565]	26.207 [35.851]	25.863 [58.974]	0.970	26.906 [44.766]	28.000 [35.518]	25.020 [57.709]	0.739	25.647 [38.903]	27.500 [34.925]	22.451 [45.148]	0.493
Number of observations	140	89	51		140	89	51		140	89	51	

Notes: Standard deviations in brackets and p-values for testing means between non-program and program facilities Significant differences over time ($p<0.05$) in the outcome variables of each study group between periods pre and post1 (marked [a]) in columns 6 and 7; between periods pre and post 2 (marked [b]) in columns 10 and 11; and between periods post1 and post2 (marked [c]) in columns 10 and 11. Significance levels at $p<0.05$.

TABLE 2A4: DESCRIPTIVE STATISTICS FOR HEALTH FACILITY RESPONDENTS

	(1) Combined mean	(2) Non-program facilities	(3) Program facilities	(5) p-value
Basic information about respondent				
1=Incharge or Deputy incharge	0.879	0.876	0.882	0.918
	[0.328]	[0.331]	[0.325]	
1=Title of medical staff (ED|PMO|CO|Nurse|Midwife)	0.793	0.787	0.804	0.807
	[0.407]	[0.412]	[0.401]	
Gender (1=Female)	0.579	0.663	0.431	0.009
	[0.496]	[0.475]	[0.500]	
Age	39.864	40.281	39.137	0.571
	[10.726]	[9.743]	[12.326]	
Religion (1=Christian)	0.907	0.899	0.922	0.649
	[0.291]	[0.303]	[0.272]	
Education (1=Attended at least Tertiary level)	0.879	0.876	0.882	0.918
	[0.328]	[0.331]	[0.325]	
Marital status (1=Married/Living with partner)	0.771	0.719	0.863	0.037
	[0.421]	[0.452]	[0.348]	
Number of years working at facility	7.521	7.449	7.647	0.870
	[6.539]	[6.152]	[7.227]	
Distance from home to facility (km)	4.884	6.200	2.588	0.063
	[13.388]	[15.912]	[6.645]	
Residential area (1=Lived in same residence since birth)	0.150	0.135	0.176	0.524
	[0.358]	[0.343]	[0.385]	
Years lived in current residence (if not lived there since birth)	8.655	8.987	8.048	0.517
	[7.998]	[8.557]	[6.914]	
Number of observations	140	89	51	

Notes: The outcomes are from URHVP Health Facility Survey. Standard deviations in brackets and t-statistics for testing means between non-program and program facilities. Staff tile/function at facility ED-Executive Director, PMO-Principal Medical Officer, CO-Clinical Officer, Nurse -Enrolled Nurse or Registered Nurse and Midwife-Enrolled midwife or Registered midwife.

TABLE 2A5: DESCRIPTIVE STATISTICS FOR WOMEN (BY OCCUPATION, ASSETS AND ACCESS TO FINANCIAL SERVICES)

	Pre-Program				Program Implementation period (post)				Program withdrawal period (post2)			
Outcomes	All (1)	Non-Program (2)	Program (3)	p-value (4)	All (5)	Non-Program (6)	Program (7)	p-value (8)	All (9)	Non-Program (10)	Program (11)	p-value (12)
Occupation												
1=Subsistence farming as woman's primary source of income	0.665 [0.473]	0.629 [0.484]	0.732 [0.445]	0.078	0.662 [0.474]	0.640 [0.481]	0.701 [0.460]	0.306	0.669 [0.471]	0.646 [0.480]	0.711 [0.455]	0.266
1=Subsistence farming as spouse's primary source of income	0.500 [0.502]	0.422 [0.496]	0.638 [0.484]	0.008	0.432 [0.496]	0.351 [0.479]	0.575 [0.497]	0.001	0.409 [0.493]	0.356 [0.480]	0.500 [0.503]	0.026
Number of assets owned per annum												
Agricultural land (acres)	1.059 [1.631]	1.131 [1.789]	0.926 [1.290]	0.275	1.286 [1.579]	1.384 [1.663]	1.106 [1.402]	0.144	1.398 [1.516]	1.524[b] [1.571]	1.165 [1.389]	0.052
Cattle	0.385 [1.773]	0.511 [2.169]	0.155 [0.464]	0.036	0.465 [1.615]	0.590 [1.882]	0.237 [0.910]	0.037	0.484 [1.166]	0.556 [1.293]	0.351 [0.878]	0.120
Goats/sheep/pigs	0.956 [2.216]	1.146 [2.507]	0.608 [1.497]	0.027	1.553 [3.394]	1.753 [3.924]	1.186[a] [2.068]	0.118	1.527 [2.487]	1.612 [2.702]	1.371[b] [2.038]	0.406
Poultry	3.887 [7.607]	4.511 [8.678]	2.729 [4.889]	0.031	6.106 [10.229]	7.084[a] [11.827]	4.292[a] [5.908]	0.010	4.491 [8.089]	5.421 [9.519]	2.783[c] [3.887]	0.001
TVs	0.051 [0.220]	0.056 [0.231]	0.041 [0.200]	0.576	0.095 [0.293]	0.129[a] [0.336]	0.031 [0.174]	0.002	0.189 [0.392]	0.230[bc] [0.422]	0.113[c] [0.319]	0.010
Radio	0.564 [0.504]	0.590 [0.505]	0.515 [0.502]	0.243	0.724 [0.464]	0.736[a] [0.455]	0.701[a] [0.482]	0.559	0.687 [0.472]	0.691[b] [0.463]	0.680[b] [0.490]	0.862
Car	0.004 [0.060]	0.000 [0.000]	0.010 [0.102]	0.320	0.004 [0.060]	0.000 [0.000]	0.010 [0.102]	0.320	0.011 [0.104]	0.011 [0.106]	0.010 [0.102]	0.943
Motorcycle	0.073 [0.260]	0.096 [0.295]	0.031 [0.174]	0.023	0.149 [0.396]	0.208[a] [0.459]	0.041 [0.200]	0.000	0.196 [0.558]	0.253[b] [0.654]	0.093 [0.292]	0.006
Bicycle	0.247 [0.432]	0.219 [0.415]	0.299 [0.460]	0.156	0.331 [0.479]	0.270 [0.445]	0.443[a] [0.520]	0.006	0.284 [0.467]	0.236 [0.439]	0.371 [0.507]	0.028
Mobile phone	0.898 [0.795]	1.017 [0.799]	0.680 [0.744]	0.000	1.247 [1.126]	1.331[a] [1.283]	1.093[a] [0.737]	0.051	1.265 [0.719]	1.303[b] [0.735]	1.196[b] [0.687]	0.228
House	0.655 [0.561]	0.674 [0.588]	0.619 [0.509]	0.414	0.756 [0.569]	0.775 [0.597]	0.722 [0.515]	0.437	0.876 [0.709]	0.910[b] [0.790]	0.814[b] [0.527]	0.232
Woman's access to financial services												
1=Owned mobile phone	0.389 [0.488]	0.478 [0.501]	0.227 [0.421]	0.000	0.582 [0.494]	0.612[a] [0.486]	0.526[a] [0.502]	0.169	0.676 [0.469]	0.691[b] [0.463]	0.649[b] [0.480]	0.488
1=Have a mobile money account	0.495 [0.502]	0.529 [0.502]	0.364 [0.492]	0.170	0.725 [0.448]	0.771[a] [0.422]	0.627[a] [0.488]	0.075	0.930 [0.256]	0.919[bc] [0.274]	0.952[bc] [0.215]	0.360
Distance to nearest mobile money agent	1.988 [2.589]	2.126 [2.716]	1.231 [1.656]	0.230	2.082 [3.033]	2.206 [3.333]	1.754 [2.046]	0.388	1.583 [2.223]	1.558 [2.094]	1.631 [2.475]	0.853
1=Have a bank account	0.022 [0.146]	0.028 [0.166]	0.010 [0.102]	0.272	0.051 [0.220]	0.062 [0.241]	0.031 [0.174]	0.223	0.087 [0.283]	0.090[b] [0.287]	0.082[b] [0.277]	0.834
Distance to nearest financial institution	15.544 [13.749]	15.671 [14.568]	15.308 [12.165]	0.827	15.161 [13.790]	15.273 [14.506]	14.954 [12.431]	0.849	14.582 [13.084]	14.569 [13.464]	14.605 [12.424]	0.983
Number of observations	275	178	97		275	178	97		275	178	97	

Notes: Standard deviations in brackets and p-values for testing means between non-program and program Women. Significant differences over time ($p<0.05$) in the outcome variables of each study group between periods pre and post1 (marked [a]) in columns 6 and 7; between periods pre and post 2 (marked [b]) in columns 10 and 11; and between periods post1 and post2 (marked [c]) in columns 10 and 11. Significance levels at $p<0.05$.

TABLE 2A6: URHVP PERCEPTIONS BY HF RESPONDENTS

	Obs. (1)	Mean (2)	Std.dev (3)
URHVP outcomes			
1= Strongly Agree URHVP increased number of women taking at least 1 ANC visit	51	0.882	0.325
1= Strongly Agree URHVP increased number of women completing 4 ANC visits	51	0.706	0.460
1= Strongly Agree URHVP increased deliveries in the unit	51	0.745	0.440
1= Strongly Agree URHVP increased PNC consultations	51	0.529	0.504
1= Strongly Agree URHVP increased Postpartum Family Planning take-up	51	0.471	0.504
1= Strongly Agree URHVP improved health seeking behavior for uncovered health services	51	0.471	0.504
1= Strongly Agree URHVP improved complications management at facility	51	0.569	0.500
1= Strongly Agree URHVP reduced maternal deaths	51	0.725	0.451
1= Strongly Agree URHVP reduced newborn deaths	51	0.686	0.469
1= Strongly Agree URHVP improved skills of medical personnel through training	51	0.569	0.500
1= Strongly Agree URHVP improved facility equipment and maintenance	51	0.490	0.505
1= Strongly Agree URHVP increased selective/preferential treatment at facility	51	0.392	0.493
1= Strongly Agree URHVP increased congestion at facility	51	0.216	0.415
1= Strongly Agree URHVP increased staff wok overload	51	0.275	0.450
1= Strongly Agree URHVP increased pregnancies/women fertility	51	0.118	0.325

TABLE 2A7: ISSUES IN SERVICE PROVISION

	All (1)	Non-program facilities (2)	Program facilities (3)	t-stats (4)	p-value (5)
Reasons for limited service provision at facility					
1=Strongly agree it's due to limited funding	0.343	0.337	0.353	-0.188	0.851
	[0.476]	[0.475]	[0.483]		
1=Strongly agree it's due to delayed remittance of funds	0.257	0.236	0.294	-0.739	0.462
	[0.439]	[0.427]	[0.460]		
1=Strongly agree it's due to lack of equipment	0.250	0.315	0.137	2.555	0.012
	[0.435]	[0.467]	[0.348]		
1=Strongly agree it's due to frequent stockouts of medicines	0.250	0.292	0.176	1.595	0.113
	[0.435]	[0.457]	[0.385]		
1=Strongly agree it's due to staff work overload	0.214	0.281	0.098	2.868	0.005
	[0.412]	[0.452]	[0.300]		
1=Strongly agree it's due to congestion	0.179	0.169	0.196	-0.400	0.690
	[0.384]	[0.376]	[0.401]		
1=Strongly agree it's due to insecurity	0.086	0.101	0.059	0.914	0.362
	[0.281]	[0.303]	[0.238]		
1=Strongly agree it's due to lack of skilled staff	0.036	0.056	0.000	2.289	0.025
	[0.186]	[0.232]	[0.000]		
Number of observations	140	89	51		
Reasons woman did not receive proper treatment					
1=Strongly Agree it's due to frequent stockouts	0.211	0.281	0.082	4.517	0.000
	[0.409]	[0.451]	[0.277]		
1=Strongly Agree it's due lack of equipment	0.091	0.107	0.062	1.328	0.186
	[0.288]	[0.310]	[0.242]		
1=Strongly Agree it's due to long waiting time	0.062	0.079	0.031	1.777	0.077
	[0.241]	[0.270]	[0.174]		
1=Strongly Agree it's due to congestion at facility	0.033	0.039	0.021	0.909	0.364
	[0.178]	[0.195]	[0.143]		
1=Strongly Agree it's due to staff absenteeism	0.029	0.022	0.041	-0.811	0.419
	[0.168]	[0.149]	[0.200]		
1=Strongly Agree it's due to demotivated staff	0.022	0.011	0.041	-1.377	0.171
	[0.146]	[0.106]	[0.200]		
1=Strongly Agree it's due to staff workload	0.015	0.011	0.021	-0.568	0.571
	[0.120]	[0.106]	[0.143]		
Number of observations	275	178	97		

Notes: Standard deviations in brackets and t-statistics for testing means between non-program and program facilities.

TABLE 2A8: PRE-PROGRAM TREND FOR HEALTH FACILITIES AND WOMEN PREGNANCIES

	Log of number of women who take:			
Panel A	(1)	(2)	(3)	(4)
	ANC1	ANC4	Delivery	PNC
Program (Program_t)	**0.452**	**0.675**	**1.447**	**-0.120**
	(1.131)	**(1.088)**	**(2.123)**	**(1.322)**
Constant	5.702**	4.990**	4.949**	6.038**
	(0.000)	(0.000)	(0.000)	(0.000)
R-Wolf p-value(program_t)	[0.832]	[0.723]	[0.584]	[0.871]
Obs.	248.000	243.000	211.000	218.000
Facility Fixed Effects	Y	Y	Y	Y
Year Fixed Effects	Y	Y	Y	Y
District time trends	Y	Y	Y	Y

Panel B	(1)	(2)	(3)	(4)
	1 if took ANC	1 if took ANC 4+	1 if Facility Delivery	1 if took PNC
Program x post	**0.202**	**0.179**	**-0.429**	**-0.064**
	(0.374)	**(0.305)**	**(0.461)**	**(0.183)**
Constant	1.800	0.777	5.035*	0.936
	(1.526)	(1.239)	(2.194)	(1.611)
R-Wolf p-value(program_wm×post)	[0.010]	[0.010]	[0.010]	[0.010]
Obs.	219.000	216.000	228.000	225.000
Women Fixed Effects	Y	Y	Y	Y
Year Fixed Effects	Y	Y	Y	Y
Controls	Y	Y	Y	Y
District time trends	Y	Y	Y	Y

Notes: Outcomes in Panel A are from MOH Administrative data for 2014 and 2015; Outcomes in Panel B are from Women pregnancy level data for 2010-2014. In Panel A are estimates for average marginal effects for number of women who take ANC, delivery and PNC services in natural log transformation (log(Y+1) while in Panel B are estimates for average marginal effects for likelihood of women taking ANC, delivery and PNC services at a health facility. All Regressions control for district linear specific time trends. In Panel B, all regressions control for woman's age at pregnancy, dummies for child sex and 1st born child, dummies for primary source of income for woman and spouse, dummies indicating possession of agricultural land, livestock, poultry, and durables (TV, radio, motorcycle, bicycle, house). Family-wise p-values (Romano-Wolf P-values), for the regressors (Program_t in Panel A and Program×Post in Panel B) are reported in square brackets (Clarke et al 2019). Standard errors clustered at the district level with health facility level fixed effects & year fixed effects (Panel A) and women fixed effects and year fixed effects (Panel B). Standard errors in parentheses. Significance levels at * $p < 0.05$, ** $p < 0.01$.

TABLE 2A9: IMPACT OF URHVP ON QUALITY OF SERVICES OF HEALTH FACILITIES (NUMBER OF BEDS AND STAFF)

Using ratios of number of staff to number of beds

	(1) Doctors/ Medical General beds	(2) Nurses/ Maternity beds	(3) Midwives/ Delivery beds	(4) Doctors/ Medical General beds	(5) Nurses/ Maternity beds	(6) Midwives/ Delivery beds
Implementation period (Post1)	-0.012	0.085	0.406**	-0.010	0.134*	0.717**
	(0.015)	(0.049)	(0.138)	(0.008)	(0.050)	(0.150)
Program × Post1	**0.036***	**0.080**	**0.124**	**0.056**	**-0.010**	**0.086**
	(0.016)	**(0.073)**	**(0.294)**	**(0.036)**	**(0.111)**	**(0.366)**
Withdrawal period (Post2)	-0.006	-0.040	0.318*	0.000	0.050	0.898**
	(0.015)	(0.088)	(0.146)	(0.000)	(0.109)	(0.121)
Program × Post2	**0.032**	**0.050**	**-0.229**	**0.070**	**-0.100**	**-0.296**
	(0.017)	**(0.112)**	**(0.180)**	**(0.054)**	**(0.217)**	**(0.242)**
Constant	0.027**	0.971**	1.834**	0.027**	0.975**	1.808**
	(0.006)	(0.030)	(0.064)	(0.002)	(0.017)	(0.044)
R-Wolf p-value (prog×post1)	[0.079]	[0.079]	[0.079]	[0.366]	[0.396]	[0.366]
R-Wolf p-value (prog×post2)	[0.089]	[0.921]	[0.782]	[0.436]	[0.861]	[0.861]
R.sq.	0.054	0.037	0.078	0.241	0.168	0.169
Obs.	266	329	325	266	329	325
Facility Fixed Effects	Y	Y	Y	Y	Y	Y
District time trends	N	N	N	Y	Y	Y
t-test (p-value)						
Post1=Prog×Post1	0.112	0.962	0.488	0.136	0.348	0.211
Post2=Prog×Post2	0.225	0.638	0.086	0.206	0.650	0.003
Prog×Post1= Prog×Post2	0.377	0.776	0.187	0.482	0.552	0.129
Control group mean in pre-program period				0.030	1.048	1.778
				(0.092)	(0.885)	(1.512)
Number of observations				42	57	56

Notes: Estimates for the ratio of number of staff to number of beds at each health facility. Regressions 4-6 control for district linear specific time trends. Family-wise p-values (Romano-Wolf P-values), for the interaction terms (Program×Post1 and Program×Post2) are reported in square brackets (Clarke et al 2019). T-test (p-value) for statistical difference between coefficients for period and the interaction term of period with program. Standard errors clustered at the district level with health facility level fixed effects. Standard errors in parentheses. Significance levels at * $p < 0.05$, ** $p < 0.01$

TABLE 2A10: IMPACT OF URHVP ON BIRTHWEIGHT

	(1) Baby weight in kgs (lowest weight used)	(2) Baby weight in kgs (highest weight used)	(3) Baby weight in kgs (lowest weight used)	(4) Baby weight in kgs (highest weight used)
Program (Program_t)	**-0.101**	**-0.185**	**-0.181**	**-0.211**
	(0.356)	**(0.367)**	**(0.252)**	**(0.330)**
Constant	3.295*	3.365**	-3.380	-3.223
	(1.301)	(1.282)	(3.930)	(3.985)
R-Wolf p-value (prog×post)	[0.960]	[0.624]	[0.990]	[0.881]
Obs.	585	585	585	585
Women Fixed Effects	Y	Y	Y	Y
Year Fixed Effects	Y	Y	Y	Y
Controls	Y	Y	Y	Y
District time trends	N	N	Y	Y
Control group mean in pre-program period			3.542	3.604
			(0.874)	(0.871)
Number of observations			346	346

Notes: Outcomes from URHVP Women Survey. Regressions control for woman's age at pregnancy, dummies for primary source of income for woman and spouse, dummies indication possession of agricultural land, livestock, poultry and durables (TV, radio, motorcycle, bicycle, house). Family-wise p-values (Romano-Wolf P-values), for the regressor (Program_t) are reported in square brackets (Clarke et al 2019). Columns 1 and 3: Missing values of birthweight are replaced by the lowest birthweight in same health facility and in the same period. Columns 2 and 4: Missing values of birthweight are replaced by the highest birthweight in same health facility and in the same period. Standard errors in parentheses. Standard errors clustered at the district level with women fixed effects and year fixed effects. Significance levels at * $p < 0.05$, ** $p < 0.01$.

TABLE 2A11: IMPACT OF URHVP ON UTILIZATION OF MATERNAL HEALTH CARE SERVICES AT HEALTH FACILITIES

Robustness check using Facility Fixed Effects Model

	Log of number of women who take:							
	(1) ANC1	(2) ANC4	(3) Delivery	(4) PNC	(5) ANC1	(6) ANC4	(7) Delivery	(8) PNC
Program (Program_t)	**0.188**	**0.321**	**0.817****	**0.590***	**0.244**	**0.382****	**0.762****	**0.562***
	(0.204)	**(0.173)**	**(0.227)**	**(0.262)**	**(0.144)**	**(0.136)**	**(0.204)**	**(0.277)**
Constant	6.954**	6.056**	6.409**	5.968**	6.302**	5.502**	5.898**	6.073**
	(0.202)	(0.167)	(0.201)	(0.234)	(0.087)	(0.083)	(0.131)	(0.154)
R-Wolf p-value (program**)**	[0.423]	[0.178]	[0.030]	[0.010]	[0.010]	[0.010]	[0.010]	[0.010]
R.sq.	0.003	0.012	0.030	0.125	0.162	0.146	0.110	0.254
Obs.	784.000	779.000	696.000	708.000	784.000	779.000	696.000	708.000
Facility Fixed Effects	Y	Y	Y	Y	Y	Y	Y	Y
Year Fixed Effects	Y	Y	Y	Y	Y	Y	Y	Y
District Time Trends	N	N	N	N	Y	Y	Y	Y

Notes: Outcomes are from MoH administrative data; estimates for average marginal effects for number of women who take ANC, delivery, and PNC services in natural log transformation (log(Y+1). Program=1 for all program facilities under the program in year t. Regressions 5-8 control for district linear specific time trends. Family-wise p-values (Romano-Wolf P-values), for the indicator variable are reported in square brackets (Clarke et al 2019). Standard errors in parentheses are clustered at the district level with health facility fixed effects and year fixed effects. Significance levels at * $p < 0.05$, ** $p < 0.01$.

TABLE 2A12: IMPACT OF URHVP ON UTILIZATION OF MATERNAL HEALTH CARE SERVICES AT HEALTH FACILITIES

Robustness check using sampling weights

	Log of number of women who take:							
Panel A	(1) ANC1	(2) ANC4	(3) Delivery	(4) PNC	(5) ANC1	(6) ANC4	(7) Delivery	(8) PNC
Program (Program_t1)	**0.406***	**0.467***	**1.067****	**0.808****	**0.528****	**0.590****	**1.187****	**1.030****
	(0.175)	**(0.182)**	**(0.224)**	**(0.260)**	**(0.158)**	**(0.173)**	**(0.253)**	**(0.359)**
Program (Program_t2)	**0.197**	**0.236**	**0.889****	**0.562***	**0.496***	**0.563***	**1.220****	**1.106**
	(0.199)	**(0.213)**	**(0.232)**	**(0.268)**	**(0.237)**	**(0.240)**	**(0.335)**	**(0.548)**
Constant	6.141**	5.157**	5.687**	5.057**	6.114**	5.147**	5.581**	4.984**
	(0.230)	(0.205)	(0.245)	(0.314)	(0.109)	(0.103)	(0.155)	(0.176)
R-Wolf p-value (program_t1)	[0.119]	[0.030]	[0.010]	[0.010]	[0.010]	[0.010]	[0.010]	[0.010]
R-Wolf p-value (program_t2)	[0.782]	[0.198]	[0.010]	[0.010]	[0.010]	[0.010]	[0.010]	[0.010]
R.sq.	0.031	0.016	0.088	0.172	0.255	0.234	0.206	0.323
Obs.	784	779	696	708	784	779	696	708
Facility Fixed Effects	Y	Y	Y	Y	Y	Y	Y	Y
Year Fixed Effects	Y	Y	Y	Y	Y	Y	Y	Y
District Time Trends	N	N	N	N	Y	Y	Y	Y
Sampling Weights	Y	Y	Y	Y	Y	Y	Y	Y
t-test (p-value)								
Program_t1= Program_t2	0.025	0.002	0.024	0.014	0.859	0.852	0.827	0.733

Panel B	(1) 1 if took ANC at health facility	(2) 1 if took ANC 4+	(3) 1 if Delivered at health facility	(4) 1 if took PNC at health facility	(5) 1 if took ANC at health facility	(6) 1 if took ANC 4+	(7) 1 if Delivered at health facility	(8) 1 if took PNC at health facility
Program (Program_t)	**0.020**	**0.024**	**0.190****	**0.091**	**0.024**	**0.039**	**0.199***	**0.069**
	(0.017)	**(0.062)**	**(0.053)**	**(0.066)**	**(0.020)**	**(0.075)**	**(0.075)**	**(0.082)**
Constant	1.070**	0.660	0.780**	0.098	1.366**	0.988	-0.539	0.372
	(0.087)	(0.509)	(0.269)	(0.247)	(0.257)	(0.650)	(0.409)	(0.341)
R-Wolf p-value (prog×post)	[0.426]	[0.852]	[0.198]	[0.198]	[0.426]	[0.426]	[0.030]	[0.426]
R.sq.	0.092	0.101	0.101	0.145	0.277	0.173	0.186	0.249
Obs.	585	581	598	591	585	581	598	591
Women fixed effects	Y	Y	Y	Y	Y	Y	Y	Y
Year Fixed Effects	Y	Y	Y	Y	Y	Y	Y	Y
District time trends	N	N	N	N	Y	Y	Y	Y
Sampling Weights	Y	Y	Y	Y	Y	Y	Y	Y

Notes: In Panel A, estimates for average marginal effects for number of women who take ANC, delivery and PNC services in natural log transformation (log(Y+1) while in Panel B, are estimates for average marginal effects for likelihood of women taking ANC, delivery and PNC services at a facility. In Panel B, all regressions control for woman's age at pregnancy, dummies for child sex and 1st born child, dummies for primary source of income for woman and spouse, dummies indicating possession of agricultural land, livestock, poultry and durables (TV, radio, motorcycle, bicycle, house). Regressions 5-8 control for district linear specific time trends. All estimates are weighted using district-level sampling weights to make our sample more representative to sample districts. Family-wise p-values (Romano-Wolf P-values), for the regressors (Program_t1 and Program_t2) in Panel A and (Program_t) in Panel B are reported in square brackets (Clarke et al 2019). T-test (p-value) for statistical difference between coefficients for program implementation and post program withdrawal for Panel A. In Panel A, standard errors clustered at the district level with health facility level fixed effects and year fixed effects, while in Panel B, standard errors clustered at the district level with women fixed effects and year fixed effects. Standard errors in parentheses. Significance levels at * $p < 0.05$, ** $p < 0.01$.

TABLE 2A13: IMPACT OF URHVP ON LIKELIHOOD OF PROVISION OF QUALITY MATERNAL HEALTH CARE SERVICES

Robustness check using sampling weights

	1 if Very Good			Daily average		1 if Very Good			Daily average	
Panel A	(1) ANC services	(2) Delivery care service	(3) PNC services	(4) Times inpatient is served	(5) Waiting time (mins)	(6) ANC services	(7) Delivery care service	(8) PNC services	(9) Times inpatient is served	(10) Waiting time (mins)
Implementation period (post)	0.099	0.093	0.112	0.670	1.190	-0.008	0.017	-0.004	0.484	0.244
	(0.058)	(0.070)	(0.060)	(0.586)	(0.740)	(0.052)	(0.053)	(0.053)	(0.598)	(0.529)
Program × Post1	**0.277***	**0.313***	**0.266***	**0.023**	**1.494**	**0.316****	**0.397***	**0.372****	**-0.161**	**1.533**
	(0.111)	**(0.141)**	**(0.107)**	**(0.701)**	**(1.637)**	**(0.106)**	**(0.156)**	**(0.098)**	**(0.744)**	**(1.984)**
Withdrawal period (Post2)	0.209*	0.202*	0.296**	0.394*	1.782*	-0.000	0.079**	0.081**	0.000	0.000
	(0.078)	(0.095)	(0.078)	(0.149)	(0.709)	(0.014)	(0.022)	(0.009)	(0.000)	(0.000)
Program × Post2	**-0.076**	**0.046**	**-0.014**	**0.206**	**0.428**	**0.002**	**0.208**	**0.194***	**-0.197**	**0.228**
	(0.126)	**(0.152)**	**(0.091)**	**(0.214)**	**(1.345)**	**(0.136)**	**(0.218)**	**(0.093)**	**(0.313)**	**(2.206)**
Constant	0.193**	0.170**	0.147**	1.263**	7.381**	0.195**	0.192**	0.158**	1.248**	7.443**
	(0.037)	(0.046)	(0.038)	(0.195)	(0.402)	(0.019)	(0.022)	(0.018)	(0.191)	(0.169)
R-Wolf. p-value (prog×post1)	[0.010]	[0.010	[0.010]	[0.139]	[0.129]	[0.040]	[0.079]	[0.030]	[0.307]	[0.307]
R-Wolf p-value (prog×post2)	[0.099]	[0.040]	[0.040]	[0.040]	[0.089]	[1.000]	[0.535]	[0.446]	[0.901]	[0.990]
R.sq.	0.123	0.122	0.206	0.011	0.056	0.332	0.411	0.443	0.021	0.213
Obs.	402	339	388	395	394	402	339	388	395	394
Facility Fixed Effects	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
District Time Trends	N	N	N	N	N	Y	Y	Y	Y	Y
Sampling Weights	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
t-test (p-value)										
Post1=Prog×Post1	0.253	0.257	0.326	0.607	0.891	0.023	0.037	0.012	0.621	0.580
Post2=Prog×Post2	0.137	0.514	0.049	0.561	0.481	0.987	0.596	0.275	0.534	0.919
Prog×Post1= Prog ×Post2	0.001	0.027	0.001	0.774	0.148	0.009	0.225	0.065	0.956	0.159

Panel B	(1) 1 if CHW services are V/good	(2) 1 if HF services are V/good	(3) 1 if medical staff services are V/good	(4) 1 if CHW services are V/good	(5) 1 if HF services are V/good	(6) 1 if Medical staff services are V/good
Program (Program_t)	**0.505****	**0.380****	**0.254****	**0.389**	**0.382***	**0.374****
	(0.162)	**(0.088)**	**(0.082)**	**(0.206)**	**(0.156)**	**(0.084)**
Constant	-0.789	0.105	0.471	-0.550	-1.197	0.827
	(0.391)	(0.370)	(0.593)	(1.039)	(0.794)	(0.721)
R-Wolf p-value (program_t)	[0.020]	[0.020]	[0.020]	[0.168]	[0.158]	[0.129]
R.sq.	0.332	0.119	0.135	0.461	0.197	0.238
Obs.	292	572	519	292	572	519
Women Fixed Effects	Y	Y	Y	Y	Y	Y
Year Fixed Effects	Y	Y	Y	Y	Y	Y
District time trends	N	N	N	Y	Y	Y
Sampling Weights	Y	Y	Y	Y	Y	Y

Notes: In Panel A, Columns 1-3 & 6-8: Very good quality of service provision taking 1 for responses "Very Good" and 0 for responses "Very Poor or Poor or Fair or Good". Columns 4-5 & 9-10: Daily average of medical staff care time. Regressions 6-10 in Panel A control for district linear specific time trends. In Panel A, standard errors clustered at the district level with health facility level fixed effects. In Panel B, are estimates for average marginal effects for quality of service provision by CHWs, health facility & medical staff. In Panel B, all regressions control for woman's age at pregnancy, dummy for 1st born child, dummies for primary source of income for woman and spouse, dummies indicating possession of agricultural land, livestock, poultry and durables (TV, radio, motorcycle, bicycle, house). Regressions 4-6 in Panel B control for district linear specific time trends. Very good quality of service provision taking 1 for responses "Very Good" and 0 for responses "Very Poor or Poor or Fair or Good". All estimates are weighted using district-level sampling weights to make our sample more representative to sample districts. In Panel B, standard errors clustered at the district level with women fixed effects and year fixed effects. Family-wise p-values (Romano-Wolf P-values), for the interaction term terms (Program×Post1 and Program×Post2) for Panel A and (Program_t) for Panel B, are reported in square brackets (Clarke et al 2019). T-test (p-value) for statistical difference between coefficients for period and the interaction term of period with program in Panel A. Standard errors in parentheses. Significance levels at * $p < 0.05$, ** $p < 0.01$

TABLE 2A14: IMPACT OF URHVP ON QUALITY OF SERVICES OF HEALTH FACILITIES (NUMBER OF BEDS AND STAFF)

Robustness check using sampling weights

	Average number at health facility						Average number at health facility					
Health facility	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
fixed effects model	Medical General beds	Maternity beds	Delivery beds	M/Doctors	Nurses	Midwives	Medical General beds	Maternity beds	Delivery beds	M/Doctors	Nurses	Midwives
Implementation period (Post1)	1.622*	0.533**	0.042	-0.043	0.818*	0.555**	-3.654**	-0.268	-0.153*	-0.161*	1.060**	1.045**
	(0.651)	(0.181)	(0.064)	(0.111)	(0.314)	(0.168)	(0.983)	(0.189)	(0.061)	(0.065)	(0.184)	(0.112)
Program × Post1	**2.620**	**0.877**	**0.321**	**0.413***	**0.903***	**1.066***	**2.686**	**1.199**	**0.360**	**0.602****	**1.263****	**1.197****
	(2.235)	**(0.616)**	**(0.221)**	**(0.174)**	**(0.438)**	**(0.391)**	**(3.137)**	**(0.886)**	**(0.237)**	**(0.209)**	**(0.421)**	**(0.411)**
Withdrawal period (Post2)	5.912*	1.546**	0.353**	0.171	1.289**	0.822**	-4.519**	-0.051	-0.032	-0.056*	1.797**	1.813**
	(2.636)	(0.397)	(0.088)	(0.186)	(0.455)	(0.187)	(0.743)	(0.232)	(0.021)	(0.024)	(0.064)	(0.026)
Program × Post2	**0.725**	**1.864**	**0.244**	**0.193**	**0.308**	**0.615***	**0.194**	**2.510**	**0.321**	**0.559***	**1.026**	**0.874****
	(4.546)	**(1.684)**	**(0.157)**	**(0.220)**	**(0.609)**	**(0.233)**	**(7.429)**	**(2.324)**	**(0.206)**	**(0.244)**	**(0.637)**	**(0.261)**
Constant	9.418**	4.827**	1.231**	0.444**	6.242**	2.444**	9.457**	4.830**	1.235**	0.448**	6.258**	2.452**
	(0.900)	(0.139)	(0.040)	(0.088)	(0.222)	(0.101)	(0.369)	(0.066)	(0.022)	(0.018)	(0.059)	(0.038)
R-Wolf p-value (prog×post1)	+	[0.218]	[0.218]	[0.139]	[0.050]	[0.069]	[0.753]	[0.614]	[0.614]	[0.069]	[0.069]	[0.069]
R-Wolf p-value (prog×post2)		[0.347]	[0.238]	[0.198]	[0.050]	[0.010]	[0.970]	[0.515]	[0.426]	[0.238]	[0.238]	[0.069]
R.sq.		0.154	0.132	0.047	0.099	0.243	0.566	0.257	0.266	0.344	0.374	0.402
Obs.	415	414	414	415	414	414	415	414	414	415	414	414
Facility Fixed Effects	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
District Time Trends	N	N	N	N	N	N	Y	Y	Y	Y	Y	Y
Sampling Weights	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
t-test (p-value)												
Post1=Prog×Post1	0.703	0.631	0.265	0.093	0.905	0.321	0.054	0.131	0.052	0.007	0.705	0.739
Post2=Prog×Post2	0.444	0.872	0.624	0.956	0.349	0.601	0.569	0.325	0.130	0.030	0.280	0.003
Prog×Post1= Prog ×Post2	0.527	0.427	0.746	0.066	0.086	0.125	0.603	0.406	0.876	0.612	0.547	0.286

Notes: Estimates for the average marginal effects for the number of beds and staff at each health facility. Regressions 7-12 control for district linear specific time trends. All estimates are weighted using district-level sampling weights to make our sample more representative to sample districts. Family-wise p-values (Romano-Wolf P-values), for the interaction terms (Program×Post1 and Program×Post2) are reported in square brackets (Clarke et al 2019). T-test (p-value) for statistical difference between coefficients for period and the interaction term of period with program. Standard errors clustered at the district level with health facility level fixed effects. Standard errors in parentheses. Significance levels at * $p < 0.05$, ** $p < 0.01$

TABLE 2A15: IMPACT OF URHVP ON MATERNAL AND CHILD HEALTH OUTCOMES

Robustness check using sampling weights

Panel A: Complications and Birthweights	(1) 1 if complication during delivery	(2) 1 if complication during delivery	(3) 1 if baby was weighed	(4) Baby weight (kg)	(5) I if baby birth weight <2.5kgs	(6) 1 if complication during delivery	(7) 1 if complication during delivery	(8) 1 if baby was weighed	(9) Baby weight (kg)	(10) I if baby birth weight <2.5kgs
Program (Program_t)	**0.096**	**0.050**	**0.201****	**-0.358**	**-0.087**	**0.078**	**0.019**	**0.204***	**0.166**	**0.037**
	(0.048)	**(0.044)**	**(0.059)**	**(0.597)**	**(0.069)**	**(0.091)**	**(0.076)**	**(0.079)**	**(0.435)**	**(0.090)**
Constant	-0.205	-0.301	0.850**	3.501**	-0.177	0.647	0.606*	-0.159	-0.488	-0.653
	(0.192)	(0.154)	(0.181)	(0.905)	(0.187)	(0.385)	(0.228)	(0.476)	(2.694)	(0.558)
R-Wolf p-value (program_t)	[0.248]	[0.347]	[0.010]	[0.852]	[0.396]	[0.525]	[0.703]	[0.010]	[0.525]	[0.703]
R.sq.	0.129	0.095	0.102	0.099	0.076	0.176	0.166	0.183	0.210	0.270
Obs.	598	598	599	514	514	598	598	599	514	514
Women Fixed Effects	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Year Fixed Effects	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Controls	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
District time trends	N	N	N	N	N	Y	Y	Y	Y	Y
Sampling weights	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y

Panel B: Pregnancy Outcomes	(1) 1 if livebirth	(2) 1 if stillbirth	(3) 1 if miscarriage	(4) 1 if neonatal death	(5) 1 if livebirth	(6) 1 if stillbirth	(7) 1 if miscarriage	(8) 1 if neonatal death
Program (Program_t)	**0.005**	**0.018**	**-0.022**	**-0.001**	**-0.027**	**0.025**	**0.002**	**0.000**
	(0.044)	**(0.015)**	**(0.027)**	**(0.018)**	**(0.059)**	**(0.019)**	**(0.035)**	**(0.025)**
Constant	1.205**	0.069	-0.216*	-0.050	1.273**	0.232	-0.489	0.011
	(0.139)	(0.035)	(0.085)	(0.050)	(0.367)	(0.134)	(0.283)	(0.170)
R-Wolf p-value (program_t)	[0.604]	[0.277]	[0.099]	[0.832]	[0.891]	[0.455]	[0.743]	[0.743]
R.sq.	0.073	0.026	0.095	0.093	0.105	0.045	0.111	0.143
Obs.	621	621	621	621	621	621	621	621
Women Fixed Effects	Y	Y	Y	Y	Y	Y	Y	Y
Year Fixed Effects	Y	Y	Y	Y	Y	Y	Y	Y
Controls	Y	Y	Y	Y	Y	Y	Y	Y
District time trends	Y	Y	Y	Y	Y	Y	Y	Y
Sampling weights	Y	Y	Y	Y	Y	Y	Y	Y
Controls	Y	Y	Y	Y	Y	Y	Y	Y

Notes: Data is from URHVP Women survey for women pregnancies from 2010-2019. All estimates are weighted using district-level sampling weights to make our sample more representative to sample districts. In Panel A, Regressions (1,2,6&7) on maternal health outcomes control for woman's age at pregnancy, dummies for child sex and 1st born child, dummies for primary source of income for woman and spouse, dummies indicating possession of agricultural land, livestock, poultry and durables (TV, radio, motorcycle, bicycle, house). Regressions 2&7 estimate the incidence of complications at birth for women who had assisted and c-section delivery. Regressions (3-5&8-10) on child health outcomes control for woman's age at pregnancy, dummies for primary source of income for woman and spouse, dummies indication possession of agricultural land, livestock, poultry and durables (TV, radio, motorcycle, bicycle, house). Family-wise p-values (Romano-Wolf P-values), for the regressor (Program_t) are reported in square brackets (Clarke et al 2019). Standard errors clustered at the district level with women fixed effects and year fixed effects.

In Panel B, Regressions control for woman's age at pregnancy, dummies for primary source of income for woman and spouse, dummies indicating possession of agricultural land, livestock, poultry and durables (TV, radio, motorcycle, bicycle, house). Family-wise p-values (Romano-Wolf P-values), for the regressor (Program_t) are reported in square brackets (Clarke et al 2019). All estimates are weighted using district-level sampling weights to make our sample more representative to sample districts. Standard errors clustered at the district level with women fixed effects and year fixed effects. Standard errors in parentheses. Significance levels at * $p < 0.05$, ** $p < 0.01$

APPENDIX 3A

TABLE 3A1: CONSTRUCTION OF SAMPLES AND DATA DESCRIPTION

Panel A: Health facilities		
Number of sampled districts	30	
Non-Program districts	12	
Program districts	18	
Number of health facilities in 12 non-program districts	515	
Sampled non-program facilities (a)		66
Number of health facilities in 18 program districts	749	
Program facilities	110	
Sampled program facilities (b)		51
Non-Program facilities	639	
Sampled non-program facilities (c)		23
Total number of sampled health facilities (b+c+a)		140
Panel B: CHWs		
Number of CHWs	272	
Non-Program CHWs from Non-Program facilities		175
Program women from Program facilities		97
Total number of observations for 3 periods	816	
Number of CHWs (sub-sample)[51]	203	
Non-Program CHWs from Non-Program facilities		118
Program women from Program facilities		85
Total number of observations for 3 periods	609	
Panel C: Data Description	**Outcome variables**	**Data source**
Measures for CHW work effort	Work coverage - Number of villages and households covered by CHW. Time use – Number of days; hours per day; hours per week spent on CHW work and CHW economic activity	Health facility survey data: pre-program period (2011-2014); program period (2015-2019); after withdrawal period (6 months 2019-2020)
Measures for trade-off effects	Number of days per month for provision of incentivized health services, non-incentivized health services and likelihood of engagement in community activities (charity, volunteer and political party)	Health facility survey data: pre-program period (2011-2014); program period (2015-2019); after withdrawal period (6 months 2019-2020)

Notes: The 30 sampled districts are from 3 regions of Uganda (western, eastern and central). The 12 non-program districts are from western region (3), eastern region (3) and central region (6) while the 18 program districts are from western region (9) and eastern region (9). Out of 110 program facilities in the program districts, one out of every two program health facilities (51 program facilities) were selected and out of 1154 non-program facilities in program and non-program districts, one out of every thirteen non-program facilities were selected (66 non-program facilities from non-program districts and 23 non-program facilities from program districts). URHVP CHW Survey targeted CHWs with more than 5 years' experience at the sampled facility. Survey targeted 2 CHW members from each health facility (2*140=280) however, we had 5 less from program facilities (Kibuku HCIV, Iyolwa HCIII, Kabwohe HCIV, Allied Health Medical Center Clinic and Bugongi HCII), 3 less from non-program facilities (Family Health Resource Center Clinic, Suubi Medical Center HCIII and Bakhita HCII).

[51] Since the characteristics for CHWs marital status (married/living with partner) and education (completed at least primary level) were significantly different between Program and Non-Program CHWs, in order to have a more comparable sample, I construct a sub sample of CHWs who are married/living with partner and obtained at least primary level education (resulting into a total of 203 CHWs).

TABLE 3A2: DESCRIPTIVE STATISTICS FOR CHW OCCUPATION, ASSETS AND ACCESS TO FINANCIAL SERVICES

	Pre-Program				Program Implementation period (post1)				Program withdrawal period (post2)			
Outcomes	All (1)	Non-Program (2)	Program (3)	p-value (4)	All (5)	Non-Program (6)	Program (7)	p-value (8)	All (9)	Non-Program (10)	Program (11)	p-value (12)
Occupation												
1=Subsistence farming as CHW's primary source of income	0.743 [0.438]	0.752 [0.434]	0.727 [0.448]	0.697	0.728 [0.446]	0.726 [0.447]	0.732 [0.445]	0.912	0.702 [0.458]	0.714 [0.453]	0.680 [0.469]	0.564
1=Subsistence farming as spouse's primary source of income	0.657 [0.476]	0.700 [0.460]	0.588 [0.496]	0.136	0.612 [0.488]	0.647 [0.479]	0.557 [0.500]	0.178	0.613 [0.488]	0.679 [0.469]	0.511 [0.503]	0.013
Number of assets owned per annum (CHW's hh)												
Agricultural land (acres)	4.341 [20.619]	4.777 [25.367]	3.542 [5.242]	0.579	4.431 [18.768]	4.595 [22.945]	4.135 [6.323]	0.804	4.581 [18.753]	4.747 [22.924]	4.282 [6.329]	0.802
Cattle	1.564 [5.285]	1.567 [6.128]	1.558 [3.255]	0.989	1.702 [4.420]	1.503 [4.134]	2.062 [4.896]	0.343	1.136 [2.139]	0.966 [1.748]	1.443 [2.689]	0.118
Goats/sheep/pigs	3.550 [6.327]	3.106 [4.460]	4364 [8.752]	0.241	4.555 [6.252]	4.194[a] [4.957]	5.206 [8.070]	0.263	3.147 [3.912]	2.663[c] [3.020]	4.021 [5.048]	0.017
Poultry	18.367 [58.727]	15.418 [49.803]	23.766 [72.338]	0.369	18.235 [40.234]	17.869 [43.579]	18.897 [33.569]	0.828	13.390 [42.468]	12.131 [35.928]	15.660 [52.378]	0.556
TVs	0.106 [0.308]	0.113 [0.318]	0.091 [0.289]	0.596	0.165 [0.372]	0.143 [0.351]	0.206[a] [0.407]	0.199	0.316 [0.466]	0.303[bc] [0.461]	0.340[bc] [0.476]	0.532
Radio	0.885 [0.347]	0.865 [0.343]	0.922 [0.354]	0.254	0.908 [0.337]	0.891 [0.330]	0.938 [0.348]	0.282	0.912 [0.500]	0.857 [0.464]	1.010 [0.550]	0.021
Car	0.014 [0.117]	0.014 [0.119]	0.013 [0.114]	0.942	0.011 [0.105]	0.011 [0.107]	0.010 [0.102]	0.932	0.022 [0.147]	0.023 [0.150]	0.021 [0.143]	0.903
Motorcycle	0.083 [0.292]	0.071 [0.258]	0.104 [0.347]	0.467	0.154 [0.362]	0.160[a] [0.368]	0.144 [0.353]	0.730	0.158 [0.365]	0.154[b] [0.362]	0.165 [0.373]	0.820
Bicycle	0.665 [0.701]	0.596 [0.621]	0.792 [0.817]	0.068	0.750 [0.711]	0.674 [0.618]	0.887 [0.840]	0.031	0.680 [0.757]	0.571 [0.682]	0.876 [0.845]	0.003
Mobile phone	1.133 [1.023]	1.092 [0.716]	1.208 [1.427]	0.507	1.419 [0.829]	1.383[a] [0.793]	1.485 [0.891]	0.350	1.779 [1.039]	1.703[bc] [0.961]	1.918[bc] [1.161]	0.123
House	1.041 [0.537]	1.043 [0.533]	1.039 [0.549]	0.963	1.184 [0.736]	1.131 [0.669]	1.278[a] [0.838]	0.140	1.283 [0.722]	1.229[b] [0.656]	1.381[b] [0.822]	0.117
Standardized Asset Index												
Asset Index	0.000 [1.410]	0.008 [1.395]	-0.015 [1.444]	0.911	0.000 [1.428]	-0.039 [1.525]	0.070 [1.238]	0.526	0.000 [1.338]	-0.160 [1.395]	0.289 [1.180]	0.005

DESCRIPTIVE STATISTICS FOR OCCUPATION, ASSETS AND ACCESS TO FINANCIAL SERVICES (CONTINUED)

	Pre-Program				Program Implementation period (post1)				Program withdrawal period (post2)			
Outcomes	All (1)	Non-Program (2)	Program (3)	p-value (4)	All (5)	Non-Program (6)	Program (7)	p-value (8)	All (9)	Non-Program (10)	Program (11)	p-value (12)
CHW's access to financial services												
1=Owned mobile phone	0.728	0.749	0.691	0.316	0.934	0.949[a]	0.907[a]	0.226	0.989	0.989[bc]	0.990[bc]	0.932
	[0.446]	[0.435]	[0.465]		[0.249]	[0.222]	[0.292]		[0.105]	[0.107]	[0.102]	
1=Have a mobile money account	0.662	0.641	0.701	0.393	0.898	0.880[a]	0.932[a]	0.160	0.993	0.988[bc]	1.000[bc]	0.158
	[0.474]	[0.481]	[0.461]		[0.304]	[0.327]	[0.254]		[0.086]	[0.107]	[0.000]	
Distance to nearest mobile money agent	3.080	3.320	2.643	0.308	2.689	2.586	2.873	0.506	1.948	1.992[b]	1.868[c]	0.731
	[4.039]	[4.551]	[2.879]		[3.354]	[3.631]	[2.804]		[3.023]	[3.279]	[2.507]	
1=Have a bank account	0.191	0.149	0.268	0.025	0.313	0.274[a]	0.381	0.076	0.441	0.411[bc]	0.495[b]	0.189
	[0.394]	[0.357]	[0.445]		[0.464]	[0.447]	[0.488]		[0.497]	[0.494]	[0.503]	
Distance to nearest financial institution	15.722	16.945	13.515	0.040	14.649	15.629	12.881	0.089	14.010	14.671	12.820	0.253
	[14.307]	[15.512]	[11.578]		[13.911]	[15.191]	[11.095]		[13.799]	[14.974]	[11.353]	
Number of observations	272	175	97		272	175	97		272	175	97	

Notes: Standard deviations in brackets and p-values for testing means between non-program and program CHWs. Asset index is a sum of all weighted standardized 'n' asset variables (livestock, poultry, TV, radio, car, motorcycle, bicycle, mobile phone, agricultural land and house) for each respondent in each period. Significant differences over time ($p<0.05$) in the outcome variables of each study group between periods pre and post1 (marked [a]) in columns 6 and 7; between periods pre and post 2 (marked [b]) in columns 10 and 11; and between periods post1 and post2 (marked [c]) in columns 10 and 11. Significance levels at $p<0.05$.

TABLE 3A3: CHALLENGES OF CHWS IN EXECUTION OF THEIR DUTIES

	Full sample				Sub sample			
	Mean All (1)	Non-program facilities (2)	Program facilities (3)	p-value (4)	Mean All (1)	Non-program facilities (2)	Program facilities (3)	p-value (5)
Lack of transport	0.728 [0.446]	0.720 [0.450]	0.742 (0.440]	0.692	0.754 [0.432]	0.763 [0.427]	0.741 (0.441]	0.728
Travelling long distances	0.640 [0.481]	0.606 [0.490]	0.701 (0.460]	0.111	0.665 [0.473]	0.636 [0.483]	0.706 (0.458]	0.294
Lack of financial incentives	0.610 [0.489]	0.646 [0.480]	0.546 (0.500]	0.113	0.601 [0.491]	0.653 [0.478]	0.529 (0.502]	0.080
Lack of support from Government officials	0.463 [0.500]	0.486 [0.501]	0.423 [0.497]	0.319	0.463 [0.500]	0.492 [0.502]	0.424 [0.497]	0.340
Inadequate training and skills	0.426 [0.495]	0.429 [0.496]	0.423 (0.497]	0.925	0.443 [0.498]	0.441 [0.499]	0.447 (0.500]	0.929
Resentment from community	0.154 [0.362]	0.131 [0.339]	0.196 (0.399]	0.181	0.172 [0.379]	0.161 [0.369]	0.188 (0.393]	0.618
Lack of recognition by community authorities	0.151 [0.358]	0.154 [0.362]	0.144 [0.353]	0.826	0.167 [0.374]	0.178 [0.384]	0.153 [0.362]	0.636
Lack of support from fellow CHWs	0.136 [0.343]	0.137 [0.345]	0.134 [0.342]	0.943	0.148 [0.356]	0.161 [0.369]	0.129 [0.338]	0.528

Notes: Standard deviations in brackets and t-statistics for testing means between non-program and program CHWs. For each challenge, 1 for responses "Strongly Agree" and 0 for responses "Strongly disagree or Disagree or Neither or Agree".

TABLE 3A4: DESCRIPTIVE STATISTICS FOR PROGRAM CHWS (EXPECTATIONS ON JOINING THE URHVP)

	Full sample			Sub sample		
	Obs. (1)	Mean (2)	Std.dev (3)	Obs. (1)	Mean (2)	Std.dev (3)
1= Strongly Agree "I expected to get skills development and experience"	97	0.907	0.465	85	0.918	0.277
1= Strongly Agree "I expected to provide social support"	97	0.691	0.465	85	0.694	0.464
1= Strongly Agree "I expected to get financial gain"	97	0.464	0.501	85	0.447	0.500
1= Strongly Agree "I expected to get job/employment opportunity"	97	0.381	0.488	85	0.388	0.490
1= Strongly Agree "I expected to get political support"	97	0.134	0.342	85	0.129	0.338

Notes: For each expectation, 1 for responses "Strongly Agree" and 0 for responses "Strongly disagree or Disagree or Neither or Agree".

TABLE 3A5: HETEROGENEITY IN THE IMPACT OF URHVP ON CHW EFFORTS (COVERAGE THAT CHW USUALLY WORKS) BY CHW GENDER

	Log of number of			
	Full sample		Sub sample	
	(1) Villages covered	(2) Households covered	(3) Villages covered	(4) Households covered
During URHVP (post1)	0.447**	0.579**	0.427**	0.680**
	(0.095)	(0.179)	(0.111)	(0.219)
Female × Post1	0.004	-0.132	-0.201	-0.348
	(0.151)	(0.188)	(0.151)	(0.218)
Program × Post1	0.642*	0.785*	0.555	0.636
	(0.313)	(0.297)	(0.322)	(0.323)
Female× prog × post1	**0.423**	**0.446**	**0.671**	**0.593**
	(0.404)	**(0.405)**	**(0.494)**	**(0.470)**
Withdrawal URHVP (post2)	0.720**	0.650	0.630**	0.728
	(0.131)	(0.342)	(0.165)	(0.442)
Female × Post2	0.094	-0.133	-0.223	-0.559
	(0.183)	(0.300)	(0.196)	(0.421)
Program × Post2	-0.045	-0.219	-0.095	-0.293
	(0.313)	(0.426)	(0.323)	(0.502)
Female × prog × post2	**-0.211**	**0.008**	**0.153**	**0.356**
	(0.301)	**(0.376)**	**(0.326)**	**(0.513)**
Constant	1.007**	4.777**	1.281**	5.738**
	(0.333)	(0.795)	(0.415)	(0.884)
R-Wolf p-value (female×prog×post1)	[0.010]	[0.010]	[0.010]	[0.010]
R-Wolf p-value (female ×prog×post2)	[0.347]	[0.347]	[0.257]	[0.277]
R.sq.	0.412	0.386	0.418	0.410
Obs.	593	591	447	445
CHW Fixed Effects	Y	Y	Y	Y
Controls	Y	Y	Y	Y
District time trends	Y	Y	Y	Y
t-test (p-value)				
female×post1= female×post2	0.327	0.996	0.864	0.480
prog×post1= female×prog×post1	0.750	0.570	0.880	0.943
prog×post2= female×prog×post2	0.764	0.723	0.655	0.376
female×prog×post1=female×prog×post2	0.036	0.329	0.187	0.642

Notes: See notes for Table 3.4. Female=1. In columns 3 and 4 is the sub sample of CHWs who are married/living with a partner and have completed at least primary education. Family-wise p-values (Romano-Wolf P-values), for the interaction terms (Female×Program×Post1 and Female×Program×Post2) are reported in square brackets (Clarke et al 2019). Significance levels at * $p < 0.05$, ** $p < 0.01$.

TABLE 3A6: HETEROGENEITY IN THE IMPACT OF URHVP ON CHW EFFORTS (COVERAGE THAT CHW USUALLY WORKS) BY CHW SELECTION

	Log of number of			
	Full sample		Sub sample	
	(1) Villages covered	(2) Households covered	(3) Villages covered	(4) Households covered
During URHVP (post1)	0.311	0.281	0.075	0.014
	(0.198)	(0.319)	(0.228)	(0.382)
Popular × Post1	0.069	0.217	0.294	0.597*
	(0.185)	(0.249)	(0.215)	(0.262)
Program × Post1	0.555	0.295	0.682	0.401
	(0.426)	(0.345)	(0.437)	(0.415)
Popular × prog × post1	**0.423**	**0.948****	**0.241**	**0.697**
	(0.401)	**(0.322)**	**(0.393)**	**(0.341)**
Withdrawal URHVP (post2)	0.474*	0.064	0.151	-0.334
	(0.231)	(0.413)	(0.273)	(0.549)
Popular × Post2	0.178	0.515	0.382	0.915**
	(0.192)	(0.252)	(0.230)	(0.319)
Program × Post2	-0.107	-0.075	0.127	0.219
	(0.527)	(0.495)	(0.556)	(0.632)
Popular × prog × post2	**0.088**	**-0.029**	**-0.061**	**-0.267**
	(0.466)	**(0.310)**	**(0.487)**	**(0.351)**
Constant	0.964**	4.794**	1.188*	5.684**
	(0.343)	(0.793)	(0.434)	(0.892)
R-Wolf p-value (popular×prog×post1)	[0.010]	[0.010]	[0.010]	[0.010]
R-Wolf p-value (popular ×prog×post2)	[0.139]	[0.139]	[0.089]	[0.109]
R.sq.	0.396	0.404	0.404	0.432
Obs.	593	591	447	445
CHW Fixed Effects	Y	Y	Y	Y
Controls	Y	Y	Y	Y
District time trends	Y	Y	Y	Y
t-test (p-value)				
popular ×post1= popular ×post2	0.103	0.028	0.511	0.171
prog×post1= popular ×prog×post1	0.871	0.289	0.589	0.665
prog×post2= popular ×prog×post2	0.844	0.949	0.854	0.585
popular ×prog×post1= popular ×prog×post2	0.270	0.003	0.359	0.009

Notes: See notes for Table 3.4. Popular=1 by Community Popular Vote. In columns 3 and 4 is the sub sample of CHWs who are married/living with a partner and have completed at least primary education. Family-wise p-values (Romano-Wolf P-values), for the interaction terms (Popular×Program×Post1 and Popular×Program×Post2) are reported in square brackets (Clarke et al 2019). Significance levels at * $p < 0.05$, ** $p < 0.01$.

TABLE 3A7: HETEROGENEITY IN THE IMPACT OF URHVP ON CHW EFFORTS (COVERAGE THAT CHW USUALLY WORKS) BY ASSET INDEX (POOREST 40%)

	Log of number of			
	Full sample		Sub sample	
	(1) Villages covered	(2) Households covered	(3) Villages covered	(4) Households covered
During URHVP (post1)	0.504**	0.688**	0.489**	0.679**
	(0.083)	(0.197)	(0.103)	(0.226)
Poor × Post1	-0.183	-0.201	-0.296	-0.171
	(0.211)	(0.297)	(0.216)	(0.388)
Program × Post1	0.707**	0.675*	0.642**	0.360
	(0.185)	(0.269)	(0.231)	(0.362)
Poor × Program × Post1	**0.267**	**0.616**	**0.271**	**1.071**
	(0.454)	**(0.568)**	**(0.632)**	**(0.654)**
Withdrawal URHVP (post2)	0.797**	0.910*	0.657**	0.910
	(0.169)	(0.393)	(0.197)	(0.483)
Poor × Post2	-0.127	-0.364	-0.181	-0.567
	(0.239)	(0.427)	(0.248)	(0.569)
Program × Post2	-0.261	-0.602	-0.094	-0.555
	(0.266)	(0.517)	(0.314)	(0.636)
Poor × Program × Post2	**0.026**	**0.158**	**-0.148**	**0.250**
	(0.355)	**(0.513)**	**(0.398)**	**(0.641)**
Constant	0.851	4.883**	0.881	5.477**
	(0.466)	(1.079)	(0.504)	(0.966)
R-Wolf p-value (poor×prog×post1)	[0.030]	[0.020]	[0.139]	[0.040]
R-Wolf p-value (poor×prog×post2)	[0.921]	[0.861]	[0.723]	[0.723]
R.sq.	0.407	0.427	0.424	0.498
Obs.	500	498	377	375
CHW Fixed Effects	Y	Y	Y	Y
Controls	Y	Y	Y	Y
District time trends	Y	Y	Y	Y
t-test (p-value)				
poor×post1= poor×post2	0.529	0.520	0.394	0.318
prog×post1=poor×prog×post1	0.422	0.932	0.635	0.417
prog×post2=poor×prog×post2	0.562	0.303	0.922	0.417
poor×prog×post1=poor×prog×post2	0.481	0.358	0.326	0.159

Notes: See notes for Table 3.4. In columns 3 and 4 is the sub sample of CHWs who are married/living with a partner and have completed at least primary education. Standardized Asset Index for pre-period is a sum of all weighted standardized 'n' asset variables (livestock, poultry, TV, radio, car, motorcycle, bicycle, mobile phone, agricultural land and house). I define the dummy Poor=1 for the bottom 40% based on their value on the asset index (Filmer and Pritchett 2001). Family-wise p-values (Romano-Wolf P-values), for the interaction terms (asset×Program×Post1 and asset×Program×Post2) are reported in square brackets (Clarke et al 2019). Significance levels at * $p < 0.05$, ** $p < 0.01$.

TABLE 3A8: HETEROGENEITY IN THE IMPACT OF URHVP ON CHW EFFORTS (DAYS PER WEEK, HOURS PER DAY AND HOURS PER WEEK SPENT ON CHW WORK AND CHW'S 1ST MAJOR ACTIVITY) BY CHW GENDER

	Log of number of											
	Full sample						Sub sample					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	Days/week CHW work	Hrs/day CHW work	Hrs/week CHW work	Days/week Major activity	Hrs/day Major activity	Hrs/week Major activity	Days/week CHW work	Hrs/day CHW work	Hrs/week CHW work	Days/week Major activity	Hrs/day Major activity	Hrs/week Major activity
During URHVP (post1)	-0.047	0.035	-0.008	-0.041	-0.015	-0.039	-0.050	0.059	0.008	-0.037	0.018	-0.004
	(0.031)	(0.049)	(0.083)	(0.038)	(0.049)	(0.077)	(0.041)	(0.053)	(0.091)	(0.039)	(0.057)	(0.076)
Female × Post1	-0.081*	-0.036	-0.149	0.072	0.083	0.169	-0.036	-0.076	-0.130	0.091*	-0.023	0.102
	(0.040)	(0.079)	(0.123)	(0.063)	(0.054)	(0.090)	(0.070)	(0.096)	(0.130)	(0.041)	(0.050)	(0.069)
Program × Post1	0.481**	0.253*	0.878**	-0.037	-0.125	-0.212	0.495**	0.198	0.840**	-0.035	-0.129	-0.193
	(0.081)	(0.099)	(0.179)	(0.056)	(0.112)	(0.168)	(0.089)	(0.099)	(0.180)	(0.059)	(0.110)	(0.161)
Female × Program × Post1	**-0.036**	**-0.038**	**-0.052**	**-0.076**	**-0.172**	**-0.251**	**0.022**	**0.029**	**0.043**	**0.019**	**0.003**	**-0.016**
	(0.120)	**(0.149)**	**(0.244)**	**(0.109)**	**(0.132)**	**(0.210)**	**(0.133)**	**(0.151)**	**(0.208)**	**(0.063)**	**(0.124)**	**(0.159)**
Withdrawal URHVP (post2)	-0.009	0.023	0.020	-0.078	-0.005	-0.040	-0.012	0.039	0.034	-0.098	0.031	-0.022
	(0.042)	(0.062)	(0.086)	(0.064)	(0.057)	(0.077)	(0.057)	(0.073)	(0.109)	(0.064)	(0.069)	(0.087)
Female × Post2	-0.123	-0.038	-0.202	0.054	0.063	0.135	-0.027	-0.109	-0.160	0.045	-0.014	0.078
	(0.068)	(0.091)	(0.141)	(0.077)	(0.049)	(0.115)	(0.096)	(0.126)	(0.155)	(0.061)	(0.051)	(0.086)
Program × Post2	0.168	-0.028	0.181	0.138	0.007	0.138	0.180	-0.102	0.114	0.151	-0.000	0.176
	(0.119)	(0.116)	(0.245)	(0.085)	(0.109)	(0.159)	(0.129)	(0.124)	(0.227)	(0.101)	(0.106)	(0.174)
Female × Program × Post2	**0.102**	**0.069**	**0.205**	**-0.014**	**-0.144**	**-0.257**	**0.016**	**0.103**	**0.127**	**0.037**	**-0.048**	**-0.158**
	(0.124)	**(0.162)**	**(0.230)**	**(0.138)**	**(0.094)**	**(0.162)**	**(0.161)**	**(0.212)**	**(0.271)**	**(0.115)**	**(0.097)**	**(0.118)**
Constant	1.400**	1.600**	2.549**	1.738**	1.773**	3.093**	1.590**	1.645**	2.836**	1.474**	1.854**	2.833**
	(0.183)	(0.193)	(0.290)	(0.197)	(0.178)	(0.215)	(0.254)	(0.269)	(0.426)	(0.196)	(0.224)	(0.332)
RWolf p-value (female×prog×post1)	[0.059]	[0.069]	[0.040]	[0.297]	[0.069]	[0.069]	[0.020]	[0.099]	[0.020]	[0.980]	[0.119]	[0.129]
RWolf p-value (female×prog×post1)	[0.465]	[0.485]	[0.347]	[0.852]	[0.475]	[0.475]	[0.614]	[0.723]	[0.574]	[0.723]	[0.644]	[0.644]
R.sq.	0.283	0.246	0.370	0.185	0.153	0.200	0.386	0.391	0.512	0.280	0.219	0.304
Obs.	593	593	593	591	591	591	447	447	447	447	447	447
CHW Fixed Effects	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Controls	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
District time trends	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
t-test (p-value)												
female×post1= female×post2	0.368	0.954	0.338	0.700	0.687	0.731	0.889	0.604	0.767	0.388	0.856	0.801
prog×post1= female ×prog×post1	0.002	0.206	0.013	0.782	0.829	0.909	0.008	0.455	0.018	0.598	0.542	0.562
prog×post2=female×prog×post2	0.762	0.695	0.956	0.417	0.393	0.176	0.509	0.490	0.977	0.495	0.781	0.200
fmale×prog×post1=fmale×prog×post2	0.246	0.465	0.273	0.564	0.754	0.976	0.961	0.619	0.715	0.861	0.580	0.358

Notes: See notes for Table 3.5. Female CHW=1**.** In columns 7-12 is the sub sample of CHWs who are married/living with a partner and have completed at least primary education. Family-wise p-values (Romano Wolf P-values), for the interaction terms (Female×Program×Post1 and Female×Program×Post2) are reported in square brackets (Clarke et al 2019). Significance levels at * $p < 0.05$, ** $p < 0.01$.

TABLE 3A9: HETEROGENEITY IN THE IMPACT OF URHVP ON CHW EFFORTS (DAYS PER WEEK, HOURS PER DAY AND HOURS PER WEEK SPENT ON CHW WORK AND CHW'S 1ST MAJOR ACTIVITY) BY CHW SELECTION

	Log of number of											
	Full sample						Sub Sample					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	Days/week CHW work	Hrs/day CHW work	Hrs/week CHW work	Days/week Major activity	Hrs/day Major activity	Hrs/week Major activity	Days/week CHW work	Hrs/day CHW work	Hrs/week CHW work	Days/week Major activity	Hrs/day Major activity	Hrs/week Major activity
During URHVP (post1)	-0.214*	0.045	-0.206	-0.024	0.028	0.016	-0.284**	-0.043	-0.413**	0.069	0.028	0.080
	(0.089)	(0.064)	(0.168)	(0.064)	(0.054)	(0.056)	(0.082)	(0.070)	(0.146)	(0.046)	(0.069)	(0.074)
Popular × Post1	0.172*	-0.017	0.192	0.009	-0.027	-0.026	0.235**	0.099	0.421**	-0.090*	-0.026	-0.098
	(0.078)	(0.064)	(0.145)	(0.076)	(0.051)	(0.075)	(0.067)	(0.064)	(0.132)	(0.044)	(0.057)	(0.074)
Program × Post1	0.498*	0.165	0.804*	-0.127	-0.028	-0.199	0.539*	0.196	0.914*	-0.189*	-0.026	-0.210
	(0.219)	(0.130)	(0.365)	(0.105)	(0.097)	(0.146)	(0.224)	(0.142)	(0.375)	(0.088)	(0.097)	(0.134)
Popular × Program × Post1	**-0.029**	**0.094**	**0.084**	**0.081**	**-0.224**	**-0.145**	**0.004**	**0.031**	**0.006**	**0.223**	**-0.153**	**0.018**
	(0.262)	**(0.172)**	**(0.480)**	**(0.141)**	**(0.130)**	**(0.209)**	**(0.293)**	**(0.165)**	**(0.506)**	**(0.116)**	**(0.104)**	**(0.162)**
Withdrawal URHVP (post2)	-0.217**	0.073	-0.182	-0.138*	-0.056	-0.190*	-0.288**	-0.045	-0.432**	-0.068	-0.018	-0.105
	(0.076)	(0.094)	(0.173)	(0.055)	(0.072)	(0.079)	(0.091)	(0.118)	(0.134)	(0.062)	(0.120)	(0.116)
Popular × Post2	0.203**	-0.058	0.180	0.094	0.079	0.192*	0.265**	0.059	0.426**	-0.010	0.058	0.085
	(0.056)	(0.075)	(0.124)	(0.060)	(0.057)	(0.070)	(0.075)	(0.074)	(0.084)	(0.058)	(0.091)	(0.088)
Program × Post2	0.350	-0.201	0.222	0.237	0.118	0.307	0.376*	-0.098	0.404	0.144	0.141	0.280
	(0.176)	(0.143)	(0.300)	(0.158)	(0.110)	(0.199)	(0.176)	(0.153)	(0.290)	(0.152)	(0.137)	(0.213)
Popular × Program × Post2	**-0.152**	**0.269**	**0.104**	**-0.134**	**-0.239***	**-0.357**	**-0.174**	**0.074**	**-0.192**	**0.041**	**-0.238***	**-0.204**
	(0.227)	**(0.163)**	**(0.404)**	**(0.188)**	**(0.103)**	**(0.216)**	**(0.262)**	**(0.171)**	**(0.444)**	**(0.159)**	**(0.115)**	**(0.174)**
Constant	1.410**	1.633**	2.589**	1.773**	1.792**	3.141**	1.566**	1.652**	2.808**	1.510**	1.829**	2.820**
	(0.195)	(0.199)	(0.311)	(0.195)	(0.177)	(0.224)	(0.258)	(0.271)	(0.444)	(0.202)	(0.225)	(0.337)
RWolf p-v (popular×prog×post1)	[0.010]	[0.010]	[0.010]	[0.416]	[0.040]	[0.040]	[0.010]	[0.010]	[0.010]	[0.901]	[0.050]	[0.089]
RWolf p-v (popular×prog×post1)	[0.515]	[0.515]	[0.257]	[0.980]	[0.515]	[0.624]	[0.505]	[0.584]	[0.465]	[0.644]	[0.505]	[0.703]
R.sq.	0.287	0.248	0.366	0.196	0.166	0.209	0.412	0.392	0.529	0.287	0.234	0.305
Obs.	593	593	593	591	591	591	447	447	447	447	447	447
CHW Fixed Effects	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Controls	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
District time trends	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
t-test (p-value)												
popular×post1= popular×post2	0.586	0.281	0.862	0.132	0.015	0.023	0.623	0.567	0.949	0.093	0.158	0.040
prog×post1=popular×prog×post1	0.271	0.807	0.388	0.390	0.358	0.869	0.302	0.579	0.302	0.044	0.487	0.395
prog×post2=popular×prog×post2	0.198	0.105	0.857	0.278	0.076	0.102	0.199	0.568	0.402	0.727	0.123	0.185
pop×prog×post1=pop×prog×post2	0.249	0.164	0.927	0.044	0.847	0.136	0.137	0.676	0.334	0.047	0.367	0.183

Notes: See notes for Table 3.5. Popular=1 by Community Popular Vote. In columns 7-12 is the sub sample of CHWs who are married/living with a partner and have completed at least primary education. Family-wise p-values (Romano-Wolf P-values), for the interaction terms (Popular×Program×Post1 and Popular×Program×Post2) are reported in square brackets (Clarke et al 2019). Significance levels at * $p < 0.05$, ** $p < 0.01$.

TABLE 3A10: HETEROGENEITY IN THE IMPACT OF URHVP ON CHW EFFORTS (DAYS PER WEEK, HOURS PER DAY AND HOURS PER WEEK SPENT ON CHW WORK AND CHW'S 1ST MAJOR ACTIVITY) BY ASSET INDEX (POOREST 40%)

	Log of number of											
	Full sample						Sub sample					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	Days/week CHW work	Hrs/day CHW work	Hrs/week CHW work	Days/week Major activity	Hrs/day Major activity	Hrs/week Major activity	Days/week CHW work	Hrs/day CHW work	Hrs/week CHW work	Days/week Major activity	Hrs/day Major activity	Hrs/week Major activity
During URHVP (post1)	-0.043	0.055	0.021	-0.017	-0.038	-0.043	-0.066	0.049	-0.019	0.003	-0.037	-0.034
	(0.041)	(0.037)	(0.085)	(0.031)	(0.036)	(0.056)	(0.036)	(0.036)	(0.065)	(0.035)	(0.049)	(0.075)
Poor × Post1	-0.234**	-0.112	-0.428*	0.090	0.176	0.251	-0.189	0.083	-0.154	-0.065	0.057	0.004
	(0.080)	(0.100)	(0.169)	(0.108)	(0.119)	(0.175)	(0.119)	(0.131)	(0.270)	(0.065)	(0.095)	(0.098)
Program × Post1	0.387**	0.183	0.706**	-0.048	-0.207	-0.291	0.455**	0.165*	0.742**	-0.019	-0.094	-0.125
	(0.138)	(0.098)	(0.235)	(0.062)	(0.126)	(0.163)	(0.133)	(0.073)	(0.192)	(0.069)	(0.119)	(0.166)
Poor × Program × Post1	**0.259**	**0.208**	**0.540**	**-0.079**	**-0.069**	**-0.156**	**0.164**	**0.110**	**0.340**	**0.042**	**-0.036**	**-0.024**
	(0.170)	**(0.157)**	**(0.304)**	**(0.129)**	**(0.151)**	**(0.208)**	**(0.195)**	**(0.165)**	**(0.347)**	**(0.103)**	**(0.119)**	**(0.152)**
Withdrawal URHVP (post2)	0.004	0.092	0.114	-0.053	-0.046	-0.082	0.003	0.071	0.089	-0.019	-0.058	-0.064
	(0.065)	(0.053)	(0.128)	(0.058)	(0.047)	(0.082)	(0.056)	(0.059)	(0.110)	(0.045)	(0.057)	(0.080)
Poor × Post2	-0.299**	-0.190*	-0.596**	0.154	0.125	0.285	-0.286	0.016	-0.354	-0.114	0.077	-0.019
	(0.099)	(0.082)	(0.151)	(0.112)	(0.135)	(0.193)	(0.164)	(0.117)	(0.259)	(0.082)	(0.086)	(0.122)
Program × Post2	0.112	-0.038	0.093	0.151	-0.041	0.062	0.068	-0.078	-0.012	0.122	0.059	0.184
	(0.119)	(0.087)	(0.220)	(0.110)	(0.126)	(0.169)	(0.122)	(0.078)	(0.204)	(0.125)	(0.133)	(0.207)
Poor × Program × Post2	**0.271**	**0.132**	**0.496**	**-0.135**	**-0.125**	**-0.251**	**0.242**	**0.043**	**0.364**	**0.065**	**-0.046**	**-0.005**
	(0.146)	**(0.122)**	**(0.248)**	**(0.157)**	**(0.142)**	**(0.210)**	**(0.205)**	**(0.125)**	**(0.306)**	**(0.120)**	**(0.082)**	**(0.125)**
Constant	1.224**	1.513**	2.257**	1.840**	1.712**	3.049**	1.535**	1.782**	2.952**	1.408**	1.699**	2.492**
	(0.255)	(0.189)	(0.339)	(0.270)	(0.237)	(0.344)	(0.269)	(0.307)	(0.485)	(0.233)	(0.226)	(0.410)
RWolf p-value (poor×prog×post1)	[0.020]	[0.020]	[0.010]	[0.307]	[0.307]	[0.099]	[0.030]	[0.020]	[0.010]	[0.436]	[0.347]	[0.198]
RWolf p-value (poor×prog×post1)	[0.990]	[1.000]	[1.000]	[1.000]	[0.990]	[1.000]	[1.000]	[1.000]	[1.000]	[1.000]	[1.000]	[1.000]
R.sq.	0.291	0.253	0.385	0.216	0.169	0.219	0.430	0.427	0.551	0.325	0.214	0.314
Obs.	500	500	500	499	499	499	377	377	377	377	377	377
Fixed Effects	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Controls	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
District time trends	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
t-test (p-value)												
poor×post1= poor×post2	0.151	0.173	0.037	0.221	0.261	0.672	0.227	0.324	0.043	0.396	0.738	0.773
prog×post1=poor×prog×post1	0.642	0.914	0.730	0.851	0.592	0.687	0.323	0.794	0.390	0.679	0.790	0.725
prog×post2=poor×prog×post2	0.419	0.298	0.220	0.212	0.714	0.339	0.516	0.470	0.344	0.769	0.555	0.473
poor×prog×post1=poor×prog×post2	0.932	0.694	0.883	0.500	0.638	0.429	0.578	0.559	0.917	0.784	0.908	0.884

Notes: See notes for Table 3.5. In columns 7-12 is the sub sample of CHWs who are married/living with a partner and have completed at least primary education. Standardized Asset Index for pre-period is a sum of all weighted standardized 'n' asset variables (livestock, poultry, TV, radio, car, motorcycle, bicycle, mobile phone, agricultural land and house). I define the dummy Poor=1 for the bottom 40% based on their value on the standardized asset index (Filmer and Pritchett 2001). Family-wise p-values (Romano-Wolf P-values), for the interaction terms (asset×Program×Post1 and asset×Program×Post2) are reported in square brackets (Clarke et al 2019). Significance levels at * $p < 0.05$, ** $p < 0.01$.

TABLE 3A11: HETEROGENEITY IN THE IMPACT OF URHVP ON CHW EFFORTS (DAYS PER MONTH SPENT ON PROVISION OF HEALTHCARE SERVICES) BY CHW GENDER

	Log of number of days per month for provision of:															
	Full sample								Sub sample							
	(1) Health Education	(2) Health campaigns	(3) Case referrals	(4) Data Collection& Reporting	(5) Women Counselling	(6) Other CHW services	(7) Treatment	(8) HIVAIDS services	(9) Health Education	(10) Health campaigns	(11) Case referrals	(12) Data Collection& Reporting	(13) Women Counselling	(14) Other CHW services	(15) Treatment	(16) HIVAIDS services
During URHVP (post1)	-0.175	-0.152*	-0.119	-0.513**	-0.169	-0.178*	-0.243	0.158	-0.128	-0.183*	-0.132	-0.487**	-0.204	-0.146	-0.377*	0.237
	(0.086)	(0.059)	(0.083)	(0.072)	(0.116)	(0.073)	(0.136)	(0.106)	(0.118)	(0.074)	(0.093)	(0.100)	(0.135)	(0.081)	(0.167)	(0.144)
Female × Post1	0.173	0.069	-0.046	0.077	0.053	-0.001	0.309	-0.052	0.200	0.055	0.037	0.044	0.290	0.072	0.379	0.075
	(0.092)	(0.121)	(0.232)	(0.093)	(0.173)	(0.061)	(0.152)	(0.162)	(0.121)	(0.134)	(0.436)	(0.084)	(0.222)	(0.096)	(0.235)	(0.284)
Program × Post1	0.383*	0.263	0.430	0.347	0.494	-0.071	0.248	0.167	0.328	0.389	0.453	0.317	0.584*	-0.036	0.403	0.115
	(0.179)	(0.174)	(0.211)	(0.196)	(0.248)	(0.159)	(0.321)	(0.341)	(0.172)	(0.191)	(0.229)	(0.203)	(0.273)	(0.164)	(0.359)	(0.375)
Female × Program × Post1	**-0.032**	**-0.215**	**-0.293**	**-0.079**	**-0.097**	**0.234**	**-0.375**	**-0.060**	**0.031**	**-0.193**	**-0.311**	**-0.168**	**-0.395**	**0.169**	**-0.391**	**-0.218**
	(0.179)	**(0.263)**	**(0.328)**	**(0.205)**	**(0.353)**	**(0.206)**	**(0.402)**	**(0.312)**	**(0.191)**	**(0.319)**	**(0.457)**	**(0.235)**	**(0.361)**	**(0.229)**	**(0.485)**	**(0.433)**
Withdrawal URHVP (post2)	-0.493**	-0.410**	-0.391**	-1.008**	-0.354	-0.377**	-0.251	0.082	-0.401*	-0.360*	-0.410*	-1.017**	-0.297	-0.325*	-0.390	0.193
	(0.145)	(0.120)	(0.121)	(0.140)	(0.192)	(0.109)	(0.175)	(0.201)	(0.166)	(0.144)	(0.171)	(0.193)	(0.240)	(0.123)	(0.206)	(0.299)
Female × Post2	0.231	-0.085	0.002	0.079	0.058	0.016	0.250	-0.033	0.365	-0.144	-0.042	0.204	0.214	0.045	0.388	0.112
	(0.138)	(0.172)	(0.283)	(0.123)	(0.188)	(0.113)	(0.264)	(0.183)	(0.180)	(0.219)	(0.522)	(0.147)	(0.221)	(0.132)	(0.428)	(0.307)
Program × Post2	-0.190	-0.179	0.317	0.238	-0.207	-0.067	0.330	-0.350	-0.233	0.068	0.384	0.240	-0.102	0.015	0.537	-0.435
	(0.268)	(0.224)	(0.233)	(0.217)	(0.274)	(0.203)	(0.341)	(0.390)	(0.247)	(0.248)	(0.258)	(0.242)	(0.331)	(0.187)	(0.398)	(0.454)
Female × Program × Post2	**0.081**	**0.353**	**-0.211**	**0.059**	**0.398**	**0.097**	**-0.378**	**-0.079**	**0.018**	**0.360**	**-0.243**	**-0.102**	**0.245**	**0.002**	**-0.459**	**-0.245**
	(0.216)	**(0.256)**	**(0.339)**	**(0.242)**	**(0.392)**	**(0.257)**	**(0.458)**	**(0.334)**	**(0.286)**	**(0.293)**	**(0.546)**	**(0.281)**	**(0.381)**	**(0.339)**	**(0.583)**	**(0.461)**
Constant	1.913**	2.027**	1.234*	0.536	1.857**	1.528**	1.049	1.815**	1.889**	2.251**	1.008*	0.957*	1.756*	1.777**	0.654	2.204**
	(0.337)	(0.414)	(0.597)	(0.435)	(0.503)	(0.327)	(0.543)	(0.463)	(0.413)	(0.380)	(0.478)	(0.370)	(0.683)	(0.409)	(0.635)	(0.545)
RWolf p-v(fmale×prog×post1)	[0.010]	[0.455]	[0.911]	[0.594]	[0.139]	[0.683]	[0.911]	[0.475]	[0.020]	[0.624]	[0.891]	[0.921]	[0.297]	[0.802]	[0.931]	[0.624]
RWolf p-v(fmale×prog×post2)	[0.347]	[0.356]	[1.000]	[0.901]	[0.267]	[1.000]	[1.000]	[1.000]	[0.396]	[0.555]	[0.990]	[0.960]	[0.396]	[0.990]	[0.990]	[0.990]
R.sq.	0.404	0.282	0.250	0.177	0.283	0.133	0.184	0.220	0.490	0.309	0.308	0.192	0.369	0.170	0.200	0.258
Obs.	593	593	593	593	593	593	593	593	447	447	447	447	447	447	447	447
CHW Fixed Effects	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Controls	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
District time trends	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
t-test (p-value)																
fmale×post1= fmale×post2	0.580	0.114	0.623	0.984	0.963	0.854	0.801	0.849	0.177	0.118	0.510	0.229	0.330	0.833	0.981	0.823
prog×post1=fmale×prog×post1	0.157	0.233	0.143	0.201	0.274	0.367	0.347	0.703	0.313	0.207	0.212	0.176	0.070	0.567	0.297	0.633
prog×post2=fmale×prog×post2	0.503	0.210	0.321	0.636	0.312	0.662	0.348	0.680	0.544	0.525	0.397	0.401	0.540	0.977	0.272	0.810
fmale×prog×post1=fmale×prog×post2	0.458	0.002	0.695	0.392	0.056	0.615	0.993	0.929	0.938	0.014	0.749	0.752	0.027	0.568	0.884	0.931

Notes: See notes for Table 3.6. Female CHW=1. In columns 9-16 is the sub sample of CHWs who are married/living with a partner and have completed at least primary education. Family-wise p-values (Romano-Wolf P-values), for the interaction terms (female×Program×Post1 and female×Program×Post2) are reported in square brackets (Clarke et al 2019). Significance levels at * $p < 0.05$, ** $p < 0.01$

TABLE 3A12: HETEROGENEITY IN THE IMPACT OF URHVP ON CHW EFFORTS (DAYS PER MONTH SPENT ON PROVISION OF HEALTHCARE SERVICES) BY CHW SELECTION

	Log of number of days per month for provision of:															
	Full sample								Sub sample							
	(1) Health Education	(2) Health campaigns	(3) Case referrals	(4) Data Collection& Reporting	(5) Women Counselling	(6) Other CHW services	(7) Treatment	(8) HIVAIDS services	(9) Health Education	(10) Health campaigns	(11) Case referrals	(12) Data Collection& Reporting	(13) Women Counselling	(14) Other CHW services	(15) Treatment	(16) HIVAIDS services
During URHVP (post1)	-0.136	-0.009	0.126	-0.302*	-0.154	-0.322	0.165	0.262	-0.176	-0.142	0.440	-0.228	0.109	-0.182	0.208	0.529
	(0.213)	(0.195)	(0.197)	(0.128)	(0.180)	(0.159)	(0.208)	(0.157)	(0.183)	(0.105)	(0.366)	(0.198)	(0.312)	(0.122)	(0.286)	(0.301)
Popular × Post1	0.036	-0.066	-0.313	-0.210	0.091	0.162	-0.367	-0.190	0.127	0.028	-0.660	-0.281	-0.171	0.028	-0.563	-0.384
	(0.174)	(0.167)	(0.183)	(0.133)	(0.109)	(0.125)	(0.220)	(0.180)	(0.128)	(0.121)	(0.357)	(0.180)	(0.275)	(0.079)	(0.291)	(0.307)
Program × Post1	0.273	-0.104	-0.296	-0.222	0.344	0.057	-0.508	-0.775	0.295	0.102	-0.659	-0.365	0.100	-0.077	-0.502	-1.120
	(0.256)	(0.350)	(0.256)	(0.303)	(0.268)	(0.186)	(0.465)	(0.537)	(0.219)	(0.293)	(0.440)	(0.330)	(0.392)	(0.128)	(0.507)	(0.626)
Popular × Program × Post1	**0.145**	**0.332**	**0.703***	**0.687***	**0.106**	**0.006**	**0.714**	**1.209***	**0.141**	**0.315**	**1.202***	**0.803***	**0.428**	**0.207**	**0.938**	**1.565***
	(0.325)	**(0.349)**	**(0.290)**	**(0.324)**	**(0.325)**	**(0.192)**	**(0.471)**	**(0.514)**	**(0.334)**	**(0.356)**	**(0.455)**	**(0.319)**	**(0.442)**	**(0.204)**	**(0.526)**	**(0.587)**
Withdrawal URHVP (post2)	-0.418	-0.324	0.019	-0.758**	-0.197	-0.482*	0.223	0.190	-0.260	-0.339	0.042	-0.595*	0.166	-0.489*	0.363	0.343
	(0.237)	(0.236)	(0.234)	(0.153)	(0.198)	(0.195)	(0.189)	(0.220)	(0.232)	(0.169)	(0.409)	(0.231)	(0.410)	(0.198)	(0.331)	(0.373)
Popular × Post2	0.021	-0.024	-0.490*	-0.250	-0.027	0.093	-0.537**	-0.216	-0.042	0.002	-0.545	-0.410*	-0.322	0.147	-0.810**	-0.249
	(0.154)	(0.186)	(0.221)	(0.145)	(0.177)	(0.134)	(0.161)	(0.174)	(0.163)	(0.155)	(0.391)	(0.162)	(0.377)	(0.180)	(0.283)	(0.307)
Program × Post2	-0.312	-0.127	-0.327	-0.264	-0.040	-0.097	-0.423	-1.182*	-0.366	0.045	-0.384	-0.452	-0.238	-0.014	-0.546	-1.363*
	(0.325)	(0.324)	(0.291)	(0.282)	(0.276)	(0.310)	(0.376)	(0.527)	(0.334)	(0.237)	(0.472)	(0.328)	(0.491)	(0.325)	(0.460)	(0.629)
Popular × Program × Post2	**0.211**	**0.099**	**0.636**	**0.686***	**-0.073**	**0.161**	**0.724**	**1.116***	**0.263**	**0.240**	**0.788**	**0.852****	**0.218**	**0.154**	**1.130***	**1.234***
	(0.406)	**(0.306)**	**(0.332)**	**(0.283)**	**(0.329)**	**(0.312)**	**(0.426)**	**(0.441)**	**(0.477)**	**(0.332)**	**(0.518)**	**(0.255)**	**(0.490)**	**(0.402)**	**(0.485)**	**(0.534)**
Constant	2.003**	2.141**	1.221	0.629	1.922**	1.521**	1.062	1.915**	1.940**	2.372**	1.153*	1.123**	1.875*	1.820**	0.725	2.472**
	(0.326)	(0.405)	(0.611)	(0.439)	(0.524)	(0.330)	(0.551)	(0.454)	(0.436)	(0.387)	(0.434)	(0.350)	(0.752)	(0.415)	(0.685)	(0.514)
RWolf p-v (pop×prog×post1)	[0.040]	[0.119]	[0.158]	[0.119]	[0.119]	[0.644]	[0.743]	[0.050]	[0.099]	[0.119]	[0.238]	[0.238]	[0.238]	[0.525]	[0.634]	[0.099]
RWolf p-v (pop×prog×post2)	[0.951]	[0.970]	[0.970]	[0.832]	[0.970]	[0.970]	[0.970]	[0.901]	[0.941]	[0.881]	[0.941]	[0.891]	[0.980]	[0.941]	[0.941]	[0.861]
R.sq.	0.396	0.258	0.261	0.190	0.269	0.133	0.189	0.245	0.479	0.290	0.333	0.204	0.353	0.167	0.209	0.295
Obs.	593	593	593	593	593	593	593	593	447	447	447	447	447	447	447	447
CHW Fixed Effects	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Controls	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
District time trends	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
t-test (p-value)																
pop×post1= pop×post2	0.908	0.673	0.218	0.638	0.506	0.405	0.186	0.751	0.390	0.842	0.407	0.194	0.527	0.435	0.215	0.423
prog×post1=pop×prog×post1	0.814	0.529	0.062	0.138	0.662	0.887	0.187	0.061	0.761	0.734	0.043	0.069	0.676	0.343	0.156	0.030
prog×post2=pop×prog×post2	0.429	0.699	0.113	0.080	0.951	0.657	0.146	0.019	0.385	0.687	0.234	0.020	0.614	0.812	0.068	0.026
pop×prog×post1=pop×prog×post2	0.789	0.313	0.727	0.996	0.570	0.598	0.964	0.663	0.699	0.755	0.041	0.812	0.560	0.884	0.456	0.237

Notes: See notes for Table 3.6. Popular=1 by Community Popular Vote. In columns 9-16 is the sub sample of CHWs who are married/living with a partner and have completed at least primary education. Family-wise p-values (Romano-Wolf P-values), for the interaction terms (Popular×Program×Post1 and Popular×Program×Post2) are reported in square brackets (Clarke et al 2019). Significance levels at * $p < 0.05$, ** $p < 0.01$

TABLE 3A13: HETEROGENEITY IN THE IMPACT OF URHVP ON CHW EFFORTS (DAYS PER MONTH SPENT ON PROVISION OF HEALTHCARE SERVICES) BY ASSET INDEX (POOREST 40%)

	Full sample								Sub sample							
Log of number of days per month for provision of:	(1) Health Education	(2) Health campaigns	(3) Case referrals	(4) Data & Reports	(5) Women Counselling	(6) Other CHW services	(7) Treatment	(8) HIVAIDS services	(9) Health Education	(10) Health campaigns	(11) Case referrals	(12) Data & Reports	(13) Women Counselling	(14) Other CHW services	(15) Treatment	(16) HIVAIDS services
During URHVP (post1)	-0.216*	-0.087	-0.261**	-0.469**	-0.120	-0.183*	-0.154	0.009	-0.197	-0.078	-0.262**	-0.455**	-0.093	-0.110	-0.208	0.053
	(0.085)	(0.076)	(0.094)	(0.070)	(0.137)	(0.068)	(0.168)	(0.135)	(0.106)	(0.085)	(0.094)	(0.083)	(0.169)	(0.080)	(0.233)	(0.161)
Poor × Post1	0.085	0.066	0.174	-0.021	0.328	-0.013	0.177	0.475	-0.046	0.046	0.424	0.161	0.407*	-0.007	0.040	0.822*
	(0.187)	(0.136)	(0.378)	(0.152)	(0.162)	(0.115)	(0.353)	(0.337)	(0.211)	(0.133)	(0.356)	(0.183)	(0.198)	(0.151)	(0.500)	(0.304)
Program × Post1	0.421**	0.263	0.364*	0.280	0.404	-0.022	0.181	0.216	0.460*	0.319	0.243	0.185	0.413	-0.039	0.331	0.056
	(0.134)	(0.152)	(0.167)	(0.188)	(0.254)	(0.111)	(0.256)	(0.267)	(0.181)	(0.204)	(0.211)	(0.192)	(0.314)	(0.138)	(0.362)	(0.338)
Poor × Program × Post1	**0.181**	**-0.312**	**-0.204**	**0.170**	**-0.319**	**0.082**	**-0.043**	**-0.396**	**0.294**	**-0.164**	**-0.055**	**0.141**	**-0.281**	**0.044**	**0.029**	**-0.546**
	(0.252)	**(0.312)**	**(0.495)**	**(0.311)**	**(0.406)**	**(0.226)**	**(0.499)**	**(0.605)**	**(0.276)**	**(0.295)**	**(0.450)**	**(0.338)**	**(0.427)**	**(0.223)**	**(0.605)**	**(0.653)**
Withdrawal URHVP (post2)	-0.556**	-0.287	-0.534**	-0.993**	-0.190	-0.416**	-0.148	-0.042	-0.475*	-0.265	-0.610**	-1.073**	-0.111	-0.341**	-0.365	0.039
	(0.169)	(0.152)	(0.135)	(0.140)	(0.238)	(0.124)	(0.187)	(0.214)	(0.191)	(0.188)	(0.141)	(0.165)	(0.308)	(0.119)	(0.241)	(0.279)
Poor × Post2	-0.015	-0.102	0.182	0.083	0.460*	0.062	0.249	0.404	-0.174	0.076	0.549	0.561**	0.617*	0.141	0.482	0.871
	(0.164)	(0.181)	(0.418)	(0.232)	(0.200)	(0.147)	(0.453)	(0.378)	(0.207)	(0.195)	(0.486)	(0.201)	(0.279)	(0.176)	(0.799)	(0.489)
Program × Post2	0.078	0.083	0.450**	0.286	-0.071	-0.179	0.267	-0.185	0.153	0.410	0.444*	0.385	0.058	-0.196	0.637	-0.270
	(0.254)	(0.196)	(0.144)	(0.240)	(0.276)	(0.172)	(0.285)	(0.403)	(0.344)	(0.274)	(0.188)	(0.211)	(0.372)	(0.164)	(0.374)	(0.493)
Poor × Program × Post2	**0.133**	**-0.365**	**-0.412**	**0.028**	**-0.710***	**0.191**	**-0.217**	**-0.604**	**0.233**	**-0.722***	**-0.610**	**-0.428**	**-0.861***	**0.110**	**-0.452**	**-1.171**
	(0.225)	**(0.308)**	**(0.461)**	**(0.280)**	**(0.317)**	**(0.259)**	**(0.635)**	**(0.599)**	**(0.232)**	**(0.306)**	**(0.570)**	**(0.284)**	**(0.347)**	**(0.268)**	**(0.905)**	**(0.702)**
Constant	1.927**	1.616*	1.157	0.562	1.602*	1.517**	1.137*	1.714*	2.169**	2.035**	1.806*	1.383**	2.025*	2.206**	1.339	2.327**
	(0.289)	(0.601)	(0.669)	(0.541)	(0.718)	(0.550)	(0.554)	(0.753)	(0.425)	(0.546)	(0.797)	(0.428)	(0.895)	(0.547)	(0.852)	(0.815)
RWolf p (poor ×prog×post1)	[0.178]	[0.960]	[0.951]	[0.782]	[0.792]	[0.951]	[0.951]	[0.871]	[0.188]	[0.941]	[0.673]	[0.634]	[0.693]	[0.891]	[0.941]	[0.782]
RWolf p(poor ×prog×post2)	[0.931]	[0.475]	[0.931]	[0.931]	[0.901]	[0.931]	[0.931]	[0.931]	[0.990]	[0.327]	[0.990]	[0.990]	[0.921]	[0.921]	[0.990]	[0.743]
R.sq.	0.398	0.276	0.259	0.205	0.277	0.189	0.207	0.235	0.479	0.352	0.358	0.241	0.360	0.243	0.211	0.314
Obs.	500	500	500	500	500	500	500	500	377	377	377	377	377	377	377	377
CHW Fixed Effects	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Controls	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
District time trends	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
t-test (p-value)																
poor×post1= poor×post2	0.457	0.104	0.923	0.460	0.315	0.494	0.722	0.629	0.352	0.798	0.676	0.002	0.118	0.221	0.269	0.860
prog×post1=poor ×prog×post1	0.318	0.169	0.336	0.798	0.229	0.724	0.737	0.407	0.554	0.276	0.596	0.924	0.294	0.790	0.688	0.476
prog×post2=poor ×prog×post2	0.865	0.243	0.117	0.537	0.212	0.276	0.565	0.606	0.841	0.014	0.112	0.031	0.089	0.405	0.296	0.353
poor×prog×post1=poor×prog×post2	0.796	0.772	0.271	0.460	0.306	0.596	0.603	0.513	0.794	0.019	0.110	0.002	0.168	0.774	0.402	0.125

Notes: See notes for Table 3.6. In columns 9-16 is the sub sample of CHWs who are married/living with a partner and have completed at least primary education. Standardized Asset Index for pre-period is a sum of all weighted standardized 'n' asset variables (agricultural land, livestock, poultry, TV, radio, car, motorcycle, bicycle). Following Filmer and Pritchett (2001), I compute the asset index (A_1) for each CHW household based on the expression: $A_{1i}=f_{11}\times(a_{1i}-a_1)/(s_1)+...+f_{1n}\times(a_{ni}-a_n)/(s_n)$. Each asset variable takes the value 1 if true, 0 otherwise. I obtain the factor weight (f_1) which is the first principal component for each variable. I then standardize each variable by subtracting the mean (a_1) of that variable from each value and dividing it by the standard deviation (s_1). Lastly, I compute the asset index for each CHW by multiplying the standardized variables by the factor weight and then add all the values together. The assets in the pre-period excludes 54 respondents who were not yet CHWs at their respective facilities. Poor=1 for the bottom 40% based on asset index. Family-wise p-values (Romano-Wolf P-values), for the interaction terms (asset×Program×Post1 and asset×Program×Post2) are reported in square brackets (Clarke et al 2019). Significance levels at * $p < 0.05$, ** $p < 0.01$.

TABLE 3A14: HETEROGENEITY IN THE IMPACT OF URHVP ON CHW'S INVOLVEMENT IN PROSOCIAL BEHAVIOR ACTIVITIES (CHARITY, VOLUNTEER AND POLITICAL PARTY) BY CHW GENDER

	Likelihood of involvement in					
	Full sample			Sub sample		
	(1) Charity	(2) Volunteer	(3) Political party	(4) Charity	(5) Volunteer	(6) Political party
During URHVP (post1)	0.255*	0.386**	0.192	0.138	0.327**	0.169
	(0.102)	(0.052)	(0.097)	(0.094)	(0.041)	(0.114)
Female × Post1	-0.057	-0.045	0.135	-0.004	0.012	0.031
	(0.119)	(0.072)	(0.155)	(0.163)	(0.092)	(0.170)
Program × Post1	0.078	0.073	0.160	0.096	0.115	0.196
	(0.257)	(0.081)	(0.232)	(0.306)	(0.076)	(0.259)
Female × prog × post1	**-0.096**	**0.136**	**-0.170**	**-0.036**	**0.116**	**-0.021**
	(0.250)	**(0.145)**	**(0.222)**	**(0.287)**	**(0.181)**	**(0.251)**
Withdrawal URHVP (post2)	0.356*	0.497**	0.188	0.189	0.415**	0.145
	(0.149)	(0.051)	(0.151)	(0.197)	(0.062)	(0.189)
Female × Post2	-0.106	0.020	0.147	-0.150	0.065	-0.032
	(0.101)	(0.088)	(0.177)	(0.158)	(0.111)	(0.185)
Program × Post2	0.078	0.212	0.188	0.095	0.265*	0.278
	(0.328)	(0.104)	(0.339)	(0.412)	(0.128)	(0.382)
Female × prog × post2	**-0.011**	**0.010**	**-0.292**	**0.101**	**0.015**	**-0.107**
	(0.260)	**(0.175)**	**(0.234)**	**(0.301)**	**(0.214)**	**(0.260)**
Constant	0.966*	0.649**	0.161	0.847*	0.504*	0.033
	(0.452)	(0.137)	(0.257)	(0.357)	(0.203)	(0.412)
R-Wolf p-value (female×prog×post1)	[0.802]	[0.079]	[0.168]	[0.574]	[0.198]	[0.337]
R-Wolf p-value (female×prog×post2)	[0.960]	[0.386]	[0.960]	[0.921]	[0.446]	[0.921]
R.sq.	0.136	0.214	0.236	0.187	0.265	0.258
Obs.	595	595	595	449	449	449
CHW Fixed Effects	Y	Y	Y	Y	Y	Y
Controls	Y	Y	Y	Y	Y	Y
District time trends	Y	Y	Y	Y	Y	Y
t-test (p-value)						
female×post1= female×post2	0.608	0.231	0.871	0.291	0.480	0.351
prog×post1=female ×prog×post1	0.718	0.743	0.437	0.807	0.996	0.637
prog×post2=female ×prog×post2	0.873	0.432	0.343	0.993	0.415	0.483
female×prog×post1=female ×prog×post2	0.433	0.262	0.370	0.353	0.492	0.551

Notes: See notes for Table 3.7. Female CHW=1. In columns 4-6 is the sub sample of CHWs who are married/living with a partner and have completed at least primary education. Family-wise p-values (Romano-Wolf P-values), for the interaction terms (Female×Program×Post1 and Female×Program×Post2) are reported in square brackets (Clarke et al 2019). Significance levels at * $p < 0.05$, ** $p < 0.01$.

TABLE 3A15: HETEROGENEITY IN THE IMPACT OF URHVP ON CHW'S INVOLVEMENT IN PROSOCIAL BEHAVIOR ACTIVITIES (CHARITY, VOLUNTEER AND POLITICAL PARTY) BY CHW SELECTION

	Likelihood of involvement in					
	Full sample			Sub sample		
	(1) Charity	(2) Volunteer	(3) Political party	(4) Charity	(5) Volunteer	(6) Political party
During URHVP (post1)	0.278*	0.209*	0.319	0.156	0.029	0.262
	(0.131)	(0.091)	(0.202)	(0.230)	(0.093)	(0.255)
Popular × Post1	-0.048	0.161	-0.127	-0.025	0.297*	-0.110
	(0.089)	(0.083)	(0.210)	(0.210)	(0.122)	(0.263)
Program × Post1	-0.028	0.119	-0.214	-0.017	0.274	-0.149
	(0.235)	(0.205)	(0.232)	(0.337)	(0.207)	(0.268)
Popular × prog × post1	**0.072**	**0.059**	**0.371**	**0.147**	**-0.039**	**0.421**
	(0.179)	**(0.244)**	**(0.229)**	**(0.284)**	**(0.269)**	**(0.271)**
Withdrawal URHVP (post2)	0.360	0.295**	0.308	0.050	0.009	0.218
	(0.202)	(0.103)	(0.271)	(0.293)	(0.134)	(0.334)
Popular × Post2	-0.044	0.210	-0.153	0.106	0.419**	-0.142
	(0.138)	(0.106)	(0.266)	(0.229)	(0.138)	(0.296)
Program × Post2	-0.000	0.259	-0.086	0.141	0.497*	0.047
	(0.248)	(0.192)	(0.370)	(0.354)	(0.219)	(0.406)
Popular × prog × post2	**0.094**	**-0.000**	**0.180**	**0.034**	**-0.136**	**0.242**
	(0.237)	**(0.229)**	**(0.303)**	**(0.320)**	**(0.234)**	**(0.315)**
Constant	0.968*	0.666**	0.149	0.894*	0.534*	0.006
	(0.459)	(0.145)	(0.283)	(0.352)	(0.240)	(0.429)
R-Wolf p-value (popular×prog×post1)	[0.327]	[0.020]	[0.079]	[0.297]	[0.069]	[0.178]
R-Wolf p-value (popular ×prog×post2)	[0.654]	[0.089]	[0.654]	[0.515]	[0.089]	[0.515]
R.sq.	0.131	0.222	0.241	0.184	0.292	0.267
Obs.	595	595	595	449	449	449
CHW Fixed Effects	Y	Y	Y	Y	Y	Y
Controls	Y	Y	Y	Y	Y	Y
District time trends	Y	Y	Y	Y	Y	Y
t-test (p-value)						
popular×post1= popular×post2	0.965	0.583	0.820	0.367	0.229	0.809
prog×post1= popular ×prog×post1	0.788	0.893	0.167	0.776	0.508	0.245
prog×post2= popular ×prog×post2	0.814	0.529	0.657	0.849	0.152	0.754
popular ×prog×post1= popular ×prog×post2	0.884	0.617	0.191	0.548	0.505	0.288

Notes: See notes for Table 3.7. Popular=1 by Community Popular Vote. In columns 4-6 is the sub sample of CHWs who are married/living with a partner and have completed at least primary education. Family-wise p-values (Romano-Wolf P-values), for the interaction terms (Popular×Program×Post1 and Popular×Program×Post2) are reported in square brackets (Clarke et al 2019).

TABLE 3A16: HETEROGENEITY IN THE IMPACT OF URHVP ON CHW'S INVOLVEMENT IN PROSOCIAL BEHAVIOR ACTIVITIES (CHARITY, VOLUNTEER AND POLITICAL PARTY) BY ASSET INDEX (POOREST 40%)

	Likelihood of involvement in					
	Full sample			Sub sample		
	(1) Charity	(2) Volunteer	(3) Political party	(4) Charity	(5) Volunteer	(6) Political party
During URHVP (post1)	0.217*	0.363**	0.153	0.128	0.328**	0.073
	(0.084)	(0.055)	(0.097)	(0.105)	(0.056)	(0.128)
Poor × Post1	0.118	0.017	-0.057	0.227*	-0.071	0.133
	(0.144)	(0.129)	(0.157)	(0.107)	(0.127)	(0.182)
Program × Post1	0.028	0.121	0.169	0.183	0.175	0.314
	(0.222)	(0.107)	(0.227)	(0.272)	(0.140)	(0.263)
Poor × Program × Post1	**-0.015**	**0.071**	**-0.068**	**-0.211**	**0.038**	**-0.204**
	(0.200)	**(0.197)**	**(0.229)**	**(0.211)**	**(0.227)**	**(0.286)**
Withdrawal URHVP (post2)	0.291	0.493**	0.089	0.141	0.432**	-0.018
	(0.162)	(0.043)	(0.166)	(0.215)	(0.068)	(0.211)
Poor × Post2	0.129	-0.027	-0.049	0.287	-0.117	0.125
	(0.160)	(0.131)	(0.196)	(0.148)	(0.159)	(0.246)
Program × Post2	0.112	0.223*	0.090	0.285	0.308*	0.315
	(0.317)	(0.093)	(0.426)	(0.394)	(0.135)	(0.494)
Poor × Program × Post2	**-0.091**	**0.080**	**0.106**	**-0.252**	**0.060**	**-0.013**
	(0.257)	**(0.214)**	**(0.242)**	**(0.292)**	**(0.236)**	**(0.300)**
Constant	0.842	0.627*	-0.101	1.105*	0.408	-0.153
	(0.451)	(0.228)	(0.311)	(0.474)	(0.348)	(0.447)
R-Wolf p-value (poor×prog×post1)	[0.446]	[0.139]	[0.446]	[0.753]	[0.753]	[0.753]
R-Wolf p-value (poor×prog×post2)	[0.733]	[0.693]	[0.723]	[0.921]	[0.921]	[0.654]
R.sq.	0.161	0.228	0.225	0.200	0.277	0.257
Obs.	500	500	500	377	377	377
CHW Fixed Effects	Y	Y	Y	Y	Y	Y
Controls	Y	Y	Y	Y	Y	Y
District time trends	Y	Y	Y	Y	Y	Y
t-test (p-value)						
poor×post1= poor×post2	0.914	0.581	0.933	0.621	0.686	0.936
prog×post1=poor×prog×post1	0.907	0.855	0.540	0.370	0.701	0.272
prog×post2=poor×prog×post2	0.690	0.612	0.978	0.387	0.468	0.639
poor×prog×post1=poor×prog×post2	0.545	0.935	0.203	0.816	0.868	0.257

Notes: See notes for Table 3.7. In columns 4-6 is the sub sample of CHWs who are married/living with a partner and have completed at least primary education. Standardized Asset Index for pre-period is a sum of all weighted standardized 'n' asset variables (livestock, poultry, TV, radio, car, motorcycle, bicycle, mobile phone, agricultural land and house). I define the dummy Poor=1 for the bottom 40% based on their value on the standardized asset index (Filmer and Pritchett 2001). Family-wise p-values (Romano-Wolf P-values), for the interaction terms (asset×Program×Post1 and asset×Program×Post2) are reported in square brackets (Clarke et al 2019). Significance levels at * $p < 0.05$, ** $p < 0.01$.

TABLE 3A17: IMPACT OF URHVP ON CHW EFFORTS (COVERAGE THAT CHW USUALLY WORKS)

Robustness check using sampling weights

	Log of number of			
	Full sample		Sub sample	
	(1) Villages covered	(2) Households covered	(3) Villages covered	(4) Households covered
During URHVP (post1)	0.407**	0.988**	0.384**	1.053**
	(0.070)	(0.115)	(0.058)	(0.145)
Program × Post1	**0.615***	**0.617***	**0.623****	**0.517**
	(0.225)	**(0.277)**	**(0.203)**	**(0.313)**
Withdrawal URHVP (post2)	0.634**	1.384**	0.651**	1.475**
	(0.074)	(0.119)	(0.106)	(0.124)
Program × Post2	**-0.368**	**-0.686**	**-0.295**	**-0.650**
	(0.296)	**(0.369)**	**(0.346)**	**(0.449)**
Constant	0.471	3.801**	0.813	5.273**
	(0.365)	(0.973)	(0.416)	(1.085)
R-Wolf p-value (prog×post1)	[0.010]	[0.010]	[0.010]	[0.010]
R-Wolf p-value (prog×post2)	[0.564]	[0.564]	[0.951]	[0.951]
R.sq.	0.388	0.360	0.375	0.397
Obs.	593	591	447	445
CHW Fixed Effects	Y	Y	Y	Y
Controls	Y	Y	Y	Y
District time trends	Y	Y	Y	Y
Sampling Weights	Y	Y	Y	Y
t-test (p-value)				
post1=program×post1	0.444	0.312	0.324	0.219
post2=program×post2	0.002	0.000	0.015	0.001
program×post1=program×post2	0.000	0.000	0.001	0.001

Notes: All outcomes are from the URHVP CHW survey. I estimate the average marginal effects for number villages and households covered by CHW in natural log transformation (log(Y+1). All regressions control for dummies for CHW's primary source of income, possession of agricultural land, livestock, poultry, durables (TV, radio, car, motorcycle, bicycle) and access to financial services (distance to nearest mobile money agent and nearest financial institution). All Regressions control for district linear specific time trends. All estimates are weighted using district-level sampling weights to make our sample more representative to sample districts. In columns 3 & 4 is the sub sample of CHWs who are married/living with a partner and have completed at least primary education. Family-wise p-values (Romano-Wolf P-values), for the interaction terms (Program×Post1 and Program×Post2) are reported in square brackets (Clarke et al 2019). T-test (p-value) for statistical difference between coefficients for period and the interaction term for period with program. Standard errors clustered at the district level with CHW fixed effects. Standard errors in parentheses. Significance levels at * $p < 0.05$, ** $p < 0.01$

TABLE 3A18: IMPACT OF URHVP ON CHW EFFORTS (DAYS PER WEEK, HOURS PER DAY AND HOURS PER WEEK SPENT ON CHW WORK AND CHW'S 1ST MAJOR ACTIVITY)

Robustness check using sampling weights

	Log of number of											
	Full sample						Sub sample					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	Days/week CHW work	Hrs/day CHW work	Hrs/week CHW work	Days/week Major activity	Hrs/day Major activity	Hrs/week Major activity	Days/week CHW work	Hrs/day CHW work	Hrs/week CHW work	Days/week Major activity	Hrs/day Major activity	Hrs/week Major activity
During URHVP (post1)	-0.076**	0.037	-0.047	-0.009	0.053	0.045	-0.052*	0.034	-0.023	0.013	0.024	0.038
	(0.018)	(0.026)	(0.042)	(0.027)	(0.029)	(0.030)	(0.023)	(0.032)	(0.056)	(0.016)	(0.025)	(0.033)
Program × Post1	**0.390****	**0.246****	**0.774****	**-0.022**	**-0.239***	**-0.289***	**0.427****	**0.244***	**0.807****	**-0.008**	**-0.134**	**-0.176**
	(0.084)	**(0.083)**	**(0.168)**	**(0.084)**	**(0.091)**	**(0.134)**	**(0.073)**	**(0.105)**	**(0.175)**	**(0.058)**	**(0.080)**	**(0.125)**
Withdrawal URHVP (post2)	-0.108**	0.018	-0.103**	-0.030	-0.005	-0.030	-0.061*	-0.017	-0.089*	-0.018	0.040	0.026
	(0.031)	(0.023)	(0.034)	(0.024)	(0.027)	(0.041)	(0.029)	(0.032)	(0.036)	(0.011)	(0.022)	(0.024)
Program × Post2	**0.125**	**0.007**	**0.170**	**0.182**	**-0.045**	**0.143**	**0.111**	**-0.025**	**0.121**	**0.213**	**0.007**	**0.234**
	(0.091)	**(0.103)**	**(0.206)**	**(0.113)**	**(0.092)**	**(0.192)**	**(0.080)**	**(0.151)**	**(0.208)**	**(0.149)**	**(0.107)**	**(0.256)**
Constant	1.302**	1.677**	2.476**	2.156**	1.855**	3.591**	1.505**	1.754**	2.824**	1.650**	1.760**	2.937**
	(0.164)	(0.164)	(0.236)	(0.209)	(0.227)	(0.252)	(0.223)	(0.279)	(0.443)	(0.185)	(0.226)	(0.331)
R-Wolf p-value (prog×post1)	[0.010]	[0.010]	[0.010]	[0.772]	[0.040]	[0.079]	[0.010]	[0.020]	[0.010]	[0.673]	[0.059]	[0.317]
R-Wolf p-value (prog×post2)	[0.584]	[0.772]	[0.564]	[0.218]	[0.792]	[0.535]	[0.753]	[0.960]	[0.812]	[0.218]	[0.960]	[0.386]
R.sq.	0.281	0.385	0.419	0.151	0.191	0.190	0.381	0.392	0.443	0.281	0.230	0.305
Obs.	593	593	593	591	591	591	447	447	447	447	447	447
CHW Fixed Effects	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Controls	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
District time trends	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Sampling weights	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
t-test (p-value)												
post1=prog×post1	0.000	0.036	0.000	0.899	0.012	0.023	0.000	0.089	0.000	0.749	0.110	0.145
post2=prog×post2	0.035	0.915	0.230	0.054	0.676	0.356	0.078	0.962	0.335	0.139	0.768	0.429
prog×post1=prog×post2	0.000	0.000	0.000	0.009	0.007	0.003	0.000	0.005	0.000	0.048	0.041	0.043

Notes: All outcomes are from the URHVP CHW survey. I estimate the average marginal effects for number days per week and hours per day spent on CHW work and CHW's major activity in natural log transformation (log(Y+1). All regressions control for dummies for primary source of income for CHW, possession of agricultural land, livestock, poultry, durables (TV, radio, car, motorcycle, bicycle) and access to financial services (distance to nearest mobile money agent and nearest financial institution). All regressions control for district linear specific time trends. All estimates are weighted using district-level sampling weights to make our sample more representative to sample districts. In columns 7-12 is the sub sample of CHWs who are married/living with a partner and have completed at least primary education. Family-wise p-values (Romano-Wolf P-values), for the interaction terms (Program×Post1 and Program×Post2) are reported in square brackets (Clarke et al 2019). T-test (p-value) for statistical difference between coefficients for period and the interaction term for period with program. Standard errors clustered at the district level with CHW fixed effects. Standard errors in parentheses. Significance levels at * $p < 0.05$, ** $p < 0.01$

TABLE 3A19: IMPACT OF URHVP ON CHW EFFORTS (DAYS PER MONTH SPENT ON PROVISION OF HEALTHCARE SERVICES)

Robustness check using sampling weights

	Log number of days per month for provision of:															
	Full sample								Sub sample							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
	Health Education	Health campaigns	Case referrals	Data Collection& Reporting	Women Counselling	Other CHW services	Treatment	HIVAIDS services	Health Education	Health campaigns	Case referrals	Data Collection& Reporting	Women Counselling	Other CHW services	Treatment	HIVAIDS services
During URHVP (post1)	-0.363**	-0.142*	-0.142*	-0.664**	-0.276*	-0.275**	-0.293	-0.020	-0.314**	-0.169*	-0.122*	-0.641**	-0.288**	-0.217**	-0.254	0.061
	(0.078)	(0.064)	(0.057)	(0.097)	(0.108)	(0.040)	(0.200)	(0.069)	(0.048)	(0.061)	(0.046)	(0.153)	(0.067)	(0.032)	(0.237)	(0.087)
Program × Post1	**0.479***	**0.048**	**0.324**	**0.311**	**0.394**	**0.082**	**0.187**	**-0.006**	**0.514****	**0.154**	**0.389***	**0.240**	**0.458***	**0.006**	**0.189**	**-0.146**
	(0.183)	**(0.131)**	**(0.192)**	**(0.197)**	**(0.194)**	**(0.133)**	**(0.351)**	**(0.268)**	**(0.182)**	**(0.137)**	**(0.184)**	**(0.265)**	**(0.211)**	**(0.166)**	**(0.392)**	**(0.337)**
Withdrawal URHVP (post2)	-0.935**	-0.605**	-0.312**	-1.347**	-0.592**	-0.521**	-0.463**	-0.254*	-0.803**	-0.614**	-0.259*	-1.354**	-0.514**	-0.445**	-0.348*	-0.203
	(0.115)	(0.033)	(0.077)	(0.070)	(0.094)	(0.077)	(0.158)	(0.094)	(0.137)	(0.033)	(0.122)	(0.075)	(0.104)	(0.053)	(0.144)	(0.145)
Program × Post2	**0.017**	**-0.171**	**0.296**	**0.333**	**-0.109**	**0.028**	**0.295**	**-0.681**	**0.061**	**0.031**	**0.339**	**0.249**	**-0.030**	**0.028**	**0.300**	**-0.984**
	(0.277)	**(0.191)**	**(0.179)**	**(0.213)**	**(0.180)**	**(0.199)**	**(0.330)**	**(0.371)**	**(0.313)**	**(0.172)**	**(0.222)**	**(0.281)**	**(0.266)**	**(0.220)**	**(0.405)**	**(0.580)**
Constant	1.434**	2.032**	1.403	0.254	1.735**	0.902**	1.756**	1.398**	1.345*	2.032**	1.147	1.080*	1.286*	1.149**	1.077	1.417**
	(0.393)	(0.337)	(0.881)	(0.503)	(0.612)	(0.302)	(0.348)	(0.483)	(0.537)	(0.335)	(0.615)	(0.412)	(0.518)	(0.414)	(0.597)	(0.439)
R-Wolf p-v(prog×post1)	[0.010]	[0.089]	[0.267]	[0.109]	[0.010]	[0.931]	[0.931]	[0.802]	[0.010]	[0.158]	[0.158]	[0.891]	[0.149]	[0.970]	[0.970]	[0.970]
R-Wolf p-v(prog×post2)	[0.980]	[0.980]	[0.426]	[0.644]	[0.792]	[0.980]	[0.970]	[0.129]	[1.000]	[1.000]	[0.663]	[1.000]	[0.842]	[1.000]	[1.000]	[0.218]
R.sq.	0.355	0.370	0.310	0.371	0.312	0.271	0.247	0.276	0.432	0.427	0.336	0.337	0.476	0.271	0.212	0.298
Obs.	593	593	593	593	593	593	593	593	447	447	447	447	447	447	447	447
CHW Fixed Effects	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Controls	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
District time trends	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Sampling weights	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
t-test (p-value)																
Post1=Program×Post1	0.001	0.311	0.036	0.001	0.020	0.019	0.371	0.966	0.001	0.082	0.019	0.033	0.006	0.187	0.468	0.607
Post2=Program×Post2	0.011	0.044	0.007	0.000	0.045	0.018	0.093	0.312	0.049	0.002	0.055	0.000	0.147	0.060	0.220	0.257
Prog×Post1= Prog×Post2	0.004	0.094	0.808	0.863	0.001	0.669	0.279	0.002	0.021	0.391	0.582	0.961	0.008	0.849	0.478	0.019

Notes: All outcomes are from the URHVP CHW survey. I estimate the average marginal effects for number days a month CHW spends on provision of various non-program health care services in natural log transformation (log(Y+1). All regressions control for district linear specific time trends. CHW Incentivized services are presented in cols 1-6 and 9-14 while CHW unincentivized services are presented in cols 7-8 and 15-16. All regressions control for dummies for primary source of income for CHW, possession of agricultural land, livestock, poultry, durables (TV, radio, car, motorcycle, bicycle) and access to financial services (distance to nearest mobile money agent and nearest financial institution). All estimates are weighted using district-level sampling weights to make our sample more representative to sample districts. In columns 9-16 is the sub sample of CHWs who are married/living with a partner and have completed at least primary education. Family-wise p-values (Romano-Wolf P-values), for the interaction terms (Program×Post1 and Program×Post2) are reported in square brackets (Clarke et al 2019). T-test (p-value) for statistical difference between coefficients for period and the interaction term for period with program.Standard errors clustered at the district level with CHW fixed effects. Standard errors in parentheses. Significance levels at * $p < 0.05$, ** $p < 0.01$

TABLE 3A20: IMPACT OF URHVP ON CHW'S INVOLVEMENT IN PROSOCIAL BEHAVIOR ACTIVITIES (CHARITY, VOLUNTEER AND POLITICAL PARTY)

Robustness check using sampling weights

	Likelihood of involvement in					
	Full sample			Sub sample		
	(1) Charity	(2) Volunteer	(3) Political party	(4) Charity	(5) Volunteer	(6) Political party
During URHVP (post1)	0.319**	0.350**	0.298**	0.296**	0.365**	0.275**
	(0.042)	(0.055)	(0.031)	(0.041)	(0.083)	(0.052)
Program × Post1	**0.022**	**0.148**	**-0.024**	**0.020**	**0.214**	**0.052**
	(0.168)	**(0.115)**	**(0.206)**	**(0.237)**	**(0.140)**	**(0.259)**
Withdrawal URHVP (post2)	0.552**	0.533**	0.365**	0.512**	0.512**	0.271**
	(0.083)	(0.058)	(0.057)	(0.122)	(0.095)	(0.071)
Program × Post2	**-0.034**	**0.149**	**-0.140**	**-0.070**	**0.290**	**0.024**
	(0.223)	**(0.183)**	**(0.308)**	**(0.333)**	**(0.217)**	**(0.392)**
Constant	1.079	0.472*	0.131	0.883*	0.209	-0.019
	(0.531)	(0.179)	(0.224)	(0.345)	(0.294)	(0.379)
R-Wolf p-value (prog×post1)	[0.327]	[0.040]	[0.099]	[0.505]	[0.020]	[0.099]
R-Wolf p-value (prog×post2)	[0.822]	[0.168]	[0.822]	[0.713]	[0.040]	[0.396]
R.sq.	0.261	0.365	0.290	0.270	0.488	0.306
Obs.	595	595	595	449	449	449
CHW Fixed Effects	Y	Y	Y	Y	Y	Y
Controls	Y	Y	Y	Y	Y	Y
District time trends	Y	Y	Y	Y	Y	Y
Sampling weights	Y	Y	Y	Y	Y	Y
t-test (p-value)						
post1=prog×post1	0.113	0.176	0.173	0.302	0.452	0.470
post2=prog×post2	0.033	0.062	0.105	0.179	0.427	0.531
prog×post1=prog×post2	0.559	0.991	0.384	0.534	0.505	0.864

Notes: All outcomes are from the URHVP CHW survey. I estimate the average marginal effects for CHWs involvement in prosocial behavior activities (Charity, Volunteer and Political party). I defined 3 dummies – Charity variable takes the value 1 if CHW was involved in blood donation, fundraising, movement for social benefits and any other charity activities; Volunteer variable takes the value 1 if CHW was involved in agricultural extension work, awareness and advocacy programs, tree planting and any other volunteer activities; and Political party variable takes the value 1 if CHW was involved in political party member campaigns, political party agent and vote counting, political party mobilization campaigns, political party communicator and other political party activities. All regressions control for dummies for CHW's primary source of income, possession of agricultural land, livestock, poultry, durables (TV, radio, car, motorcycle, bicycle) and access to financial services (distance to nearest mobile money agent and nearest financial institution). All regressions control for district linear specific time trends. All estimates are weighted using district-level sampling weights to make our sample more representative to sample districts. In columns 4-6 is the sub sample of CHWs who are married/living with a partner and have completed at least primary education. Family-wise p-values (Romano-Wolf P-values), for the interaction terms (Program×Post1 and Program×Post2) are reported in square brackets (Clarke et al 2019). T-test (p-value) for statistical difference between coefficients for period and the interaction term for period with program. Standard errors clustered at the district level with CHW fixed effects. Standard errors in parentheses. Significance levels at * $p < 0.05$, ** $p < 0.01$

TABLE 3A21: IMPACT OF URHVP ON CHW EFFORTS (COVERAGE THAT CHW USUALLY WORKS)

Robustness check excluding CHW time-variant characteristics

	Log of number of			
	(1) Villages covered	(2) Households covered	(3) Villages covered	(4) Households covered
During URHVP (post1)	0.172**	0.413**	0.212**	0.478**
	(0.027)	(0.064)	(0.041)	(0.075)
Program × Post1	**0.861****	**0.859****	**0.835****	**0.747****
	(0.101)	**(0.168)**	**(0.106)**	**(0.183)**
Withdrawal URHVP (post2)	0.236**	0.327**	0.315**	0.475**
	(0.031)	(0.075)	(0.062)	(0.125)
Program × Post2	**-0.021**	**-0.246**	**-0.081**	**-0.467**
	(0.067)	**(0.157)**	**(0.125)**	**(0.250)**
Constant	1.265**	4.888**	1.263**	4.900**
	(0.022)	(0.045)	(0.016)	(0.025)
R-Wolf p-value (prog×post1)	[0.010]	[0.010]	[0.010]	[0.010]
R-Wolf p-value (prog×post2)	[0.040]	[0.555]	[0.693]	[0.158]
R.sq.	0.339	0.282	0.357	0.322
Obs.	759	757	759	757
CHW Fixed Effects	Y	Y	Y	Y
District time trends	N	N	Y	Y
t-test (p-value)				
post1=program×post1	0.000	0.033	0.000	0.268
post2=program×post2	0.005	0.009	0.043	0.018
program×post1=program×post2	0.000	0.000	0.000	0.000

Notes: All outcomes are from the URHVP CHW survey.I estimate the average marginal effects for number villages and households covered by CHW in natural log transformation (log(Y+1) for full sample. Regressions in columns 3-4 control for district linear specific time trends. Family-wise p-values (Romano-Wolf P-values), for the interaction terms (Program×Post1 and Program×Post2 are reported in square brackets (Clarke et al 2019). T-test (p-value) for statistical difference between coefficients for period and the interaction term for period with program. Standard errors clustered at the district level with CHW fixed effects. Standard errors in parentheses. Significance levels at * $p < 0.05$, ** $p < 0.01$

TABLE 3A22: IMPACT OF URHVP ON CHW EFFORTS (DAYS PER WEEK, HOURS PER DAY AND HOURS PER WEEK SPENT ON CHW WORK AND CHW'S 1ST MAJOR ACTIVITY)

Robustness check excluding CHW time-variant characteristics

	Log of number of											
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	Days/week CHW work	Hrs/day CHW work	Hrs/week CHW work	Days/week Major activity	Hrs/day Major activity	Hrs/week Major activity	Days/week CHW work	Hrs/day CHW work	Hrs/week CHW work	Days/week Major activity	Hrs/day Major activity	Hrs/week Major activity
During URHVP	0.018	0.082**	0.110*	-0.018	-0.008	-0.043	0.077**	0.030	0.133**	0.005	-0.010	-0.001
	(0.020)	(0.022)	(0.041)	(0.012)	(0.022)	(0.028)	(0.026)	(0.016)	(0.037)	(0.016)	(0.019)	(0.025)
Program × Post1	**0.263****	**0.210****	**0.585****	**-0.117***	**-0.103****	**-0.252****	**0.257****	**0.233****	**0.600****	**-0.103**	**-0.112****	**-0.256****
	(0.056)	**(0.036)**	**(0.082)**	**(0.048)**	**(0.035)**	**(0.053)**	**(0.065)**	**(0.034)**	**(0.080)**	**(0.051)**	**(0.035)**	**(0.057)**
Withdrawal URHVP	0.033	0.108**	0.157**	-0.077**	-0.028	-0.132**	0.150**	0.003	0.203**	-0.031	-0.029	-0.047
	(0.023)	(0.035)	(0.046)	(0.016)	(0.022)	(0.038)	(0.029)	(0.021)	(0.031)	(0.024)	(0.022)	(0.034)
Program × Post2	**-0.000**	**-0.051**	**-0.051**	**0.032**	**0.049**	**0.076**	**-0.013**	**0.000**	**-0.013**	**0.062**	**0.028**	**0.069**
	(0.043)	**(0.043)**	**(0.073)**	**(0.040)**	**(0.046)**	**(0.054)**	**(0.058)**	**(0.042)**	**(0.061)**	**(0.048)**	**(0.045)**	**(0.068)**
Constant	1.311**	1.578**	2.394**	1.784**	1.696**	3.131**	1.313**	1.578**	2.396**	1.784**	1.697**	3.132**
	(0.014)	(0.014)	(0.025)	(0.012)	(0.013)	(0.017)	(0.010)	(0.006)	(0.014)	(0.006)	(0.005)	(0.009)
R-Wolf p-value (prog×post1)	[0.010]	[0.010]	[0.010]	[0.010]	[0.010]	[0.010]	[0.010]	[0.010]	[0.010]	[0.089]	[0.089]	[0.020]
R-Wolf p-value (prog×post2)	[0.673]	[0.465]	[0.406]	[0.465]	[0.673]	[0.673]	[1.000]	[1.000]	[1.000]	[0.475]	[0.941]	[0.753]
R.sq.	0.147	0.157	0.249	0.070	0.038	0.089	0.190	0.205	0.297	0.128	0.081	0.127
Obs.	759	759	759	757	757	757	759	759	759	757	757	757
CHW Fixed Effects	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
District time trends	N	N	N	N	N	N	Y	Y	Y	Y	Y	Y
t-test (p-value)												
post1=prog×post1	0.001	0.019	0.000	0.070	0.071	0.006	0.039	0.000	0.000	0.096	0.055	0.002
post2=prog×post2	0.568	0.042	0.061	0.032	0.205	0.022	0.073	0.969	0.026	0.209	0.402	0.267
prog×post1=prog×post2	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Notes: All outcomes are from the URHVP CHW survey. I estimate the average marginal effects for number days per week and hours per day spent on CHW work and CHW's major activity in natural log transformation (log(Y+1) for full sample. Regressions in Cols 7-12 control for district linear specific time trends. Family-wise p-values (Romano-Wolf P-values), for the interaction terms (Program×Post1 and Program×Post2) are reported in square brackets (Clarke et al 2019). T-test (p-value) for statistical difference between coefficients for period and the interaction term for period with program. Standard errors clustered at the district level with CHW fixed effects. Standard errors in parentheses. Significance levels at * $p < 0.05$, ** $p < 0.01$

TABLE 3A23: IMPACT OF URHVP ON CHW EFFORTS (DAYS PER MONTH SPENT ON PROVISION OF HEALTHCARE SERVICES)

Robustness check excluding CHW time-variant characteristics

	Log number of days per month for provision of:															
	(1) Health Education	(2) Health campaigns	(3) Case referrals	(4) Data Collection& Reporting	(5) Women Counselling	(6) Other CHW services	(7) Treatment	(8) HIVAIDS services	(9) Health Education	(10) Health campaigns	(11) Case referrals	(12) Data Collection& Reporting	(13) Women Counselling	(14) Other CHW services	(15) Treatment	(16) HIVAIDS services
During URHVP (post1)	0.000	0.131**	0.027	-0.016	0.054	0.040	0.119	0.096**	-0.025	0.288**	-0.012	-0.345**	-0.099	-0.164**	0.156	0.009
	(0.039)	(0.032)	(0.070)	(0.045)	(0.077)	(0.033)	(0.097)	(0.033)	(0.043)	(0.048)	(0.056)	(0.038)	(0.069)	(0.035)	(0.105)	(0.039)
Program × Post1	**0.586****	**0.302****	**0.379****	**0.249***	**0.508****	**0.040**	**-0.081**	**0.360***	**0.570****	**0.353****	**0.403****	**0.230***	**0.447***	**0.005**	**0.006**	**0.213**
	(0.078)	**(0.083)**	**(0.131)**	**(0.094)**	**(0.137)**	**(0.080)**	**(0.176)**	**(0.148)**	**(0.088)**	**(0.127)**	**(0.135)**	**(0.092)**	**(0.165)**	**(0.078)**	**(0.208)**	**(0.123)**
Withdrawal URHVP (post2)	-0.171**	0.055	-0.020	-0.031	0.031	0.001	0.008	0.038	-0.235**	0.372**	-0.116	-0.696**	-0.292**	-0.420**	0.099	-0.154*
	(0.051)	(0.048)	(0.074)	(0.044)	(0.078)	(0.052)	(0.110)	(0.061)	(0.067)	(0.097)	(0.065)	(0.062)	(0.076)	(0.057)	(0.121)	(0.058)
Program × Post2	**0.125**	**0.085**	**0.150**	**0.088**	**-0.041**	**-0.043**	**-0.041**	**-0.051**	**0.123**	**0.161**	**0.239**	**0.077**	**-0.125**	**-0.088**	**0.110**	**-0.274***
	(0.088)	**(0.092)**	**(0.134)**	**(0.069)**	**(0.108)**	**(0.088)**	**(0.161)**	**(0.145)**	**(0.133)**	**(0.194)**	**(0.130)**	**(0.123)**	**(0.152)**	**(0.114)**	**(0.242)**	**(0.115)**
Constant	1.738**	1.392**	1.880**	0.817**	1.658**	1.617**	0.970**	1.407**	1.737**	1.388**	1.877**	0.820**	1.655**	1.613**	0.976**	1.412**
	(0.025)	(0.024)	(0.043)	(0.025)	(0.041)	(0.028)	(0.050)	(0.040)	(0.011)	(0.011)	(0.018)	(0.016)	(0.020)	(0.009)	(0.020)	(0.016)
R-Wolf p-v(prog×post1)	[0.010]	[0.010]	[0.010]	[0.010]	[0.010]	[0.495]	[0.822]	[0.010]	[0.010]	[0.010]	[0.030]	[0.089]	[0.020]	[0.703]	[0.703]	[0.059]
R-Wolf p-v(prog×post2)	[0.960]	[0.624]	[0.782]	[0.960]	[0.990]	[0.960]	[0.990]	[0.990]	[0.753]	[0.594]	[0.594]	[0.941]	[0.713]	[0.703]	[0.782]	[0.545]
R.sq.	0.210	0.090	0.046	0.033	0.125	0.011	0.006	0.084	0.255	0.153	0.138	0.098	0.212	0.107	0.077	0.185
Obs.	759.000	759.000	759.000	759.000	759.000	759.000	759.000	759.000	759.000	759.000	759.000	759.000	759.000	759.000	759.000	759.000
CHW Fixed Effects	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
District time trends	N	N	N	N	N	N	N	N	Y	Y	Y	Y	Y	Y	Y	Y
t-test (p-value)																
Post1=Program×Post1	0.000	0.106	0.057	0.036	0.027	0.996	0.435	0.111	0.000	0.703	0.027	0.000	0.020	0.128	0.627	0.183
Post2=Program×Post2	0.023	0.803	0.352	0.246	0.661	0.713	0.849	0.613	0.084	0.475	0.079	0.000	0.470	0.061	0.975	0.494
Prog×Post1= Prog×Post2	0.000	0.018	0.046	0.073	0.000	0.141	0.678	0.001	0.000	0.101	0.133	0.196	0.000	0.222	0.378	0.000

Notes: All outcomes are from the URHVP CHW survey. I estimate the average marginal effects for number days a month CHW spends on provision of various health care services in natural log transformation (log(Y+1) for full sample. Outcomes in Cols 9-16 control for district linear specific time trends. CHW Incentivized services are presented in cols 1-6 and 9-14 while CHW unincentivized services are presented in cols 7-8 and 15-16. Family-wise p-values (Romano-Wolf P-values), for the interaction terms (Program×Post1 and Program×Post2) respectively are reported in square brackets (Clarke et al 2019). T-test (p-value) for statistical difference between coefficients for period and the interaction term for period with program. Standard errors clustered at the district level with CHW fixed effects. Standard errors in parentheses. Significance levels at * $p < 0.05$, ** $p < 0.01$

TABLE 3A24: IMPACT OF URHVP ON CHW'S INVOLVEMENT IN PROSOCIAL BEHAVIOR ACTIVITIES (CHARITY, VOLUNTEER AND POLITICAL PARTY)

Robustness check excluding CHW time-variant characteristics

	(1) Charity	(2) Volunteer	(3) Political party	(4) Charity	(5) Volunteer	(6) Political party
During URHVP (post1)	0.168**	0.148**	0.077	0.247**	0.309**	0.131*
	(0.039)	(0.033)	(0.038)	(0.043)	(0.040)	(0.059)
Program × Post1	**-0.029**	**0.136***	**0.079**	**-0.052**	**0.162***	**0.036**
	(0.062)	**(0.057)**	**(0.051)**	**(0.090)**	**(0.079)**	**(0.087)**
Withdrawal URHVP (post2)	0.116**	0.079*	-0.134**	0.276**	0.405**	-0.026
	(0.028)	(0.030)	(0.031)	(0.060)	(0.055)	(0.077)
Program × Post2	**-0.008**	**0.143***	**0.146***	**-0.053**	**0.189**	**0.052**
	(0.057)	**(0.055)**	**(0.060)**	**(0.120)**	**(0.110)**	**(0.154)**
Constant	0.575**	0.659**	0.240**	0.577**	0.661**	0.239**
	(0.017)	(0.017)	(0.013)	(0.010)	(0.009)	(0.012)
R-Wolf p-value (prog×post1)	[0.010]	[0.010]	[0.010]	[0.356]	[0.010]	[0.059]
R-Wolf p-value (prog×post1)	[0.089]	[0.010]	[0.832]	[0.921]	[0.218]	[0.921]
R.sq.	0.057	0.107	0.107	0.097	0.161	0.149
Obs.	762	762	762	762	762	762
CHW Fixed Effects	Y	Y	Y	Y	Y	Y
District time trends	N	N	N	Y	Y	Y
t-test (p-value)						
post1=prog×post1	0.045	0.890	0.982	0.029	0.207	0.515
post2=prog×post2	0.107	0.402	0.002	0.077	0.200	0.738
prog×post1=prog×post2	0.683	0.855	0.329	0.990	0.616	0.874

Notes: All outcomes are from the URHVP CHW survey. I estimate the average marginal effects for CHW's likelihood of involvement in prosocial behavior activities (Charity, Volunteer and Political party) for full sample. I defined 3 dummies – Charity variable takes the value 1 if CHW was involved in blood donation, fundraising, movement for social benefits and any other charity activities; Volunteer variable takes the value 1 if CHW was involved in agricultural extension work, awareness and advocacy programs, tree planting and any other volunteer activities; and Political party variable takes the value 1 if CHW was involved in political party member campaigns, political party agent and vote counting, political party mobilization campaigns, political party communicator and other political party activities. Regressions in columns 4-6 control for district linear specific time trends. Family-wise p-values (Romano-Wolf P-values), for the interaction terms (Program×Post1 and Program×Post2) are reported in square brackets (Clarke et al 2019). T-test (p-value) for statistical difference between coefficients for period and the interaction term for period with program. Standard errors clustered at the district level with CHW fixed effects. Standard errors in parentheses. Significance levels at * $p < 0.05$, ** $p < 0.01$

TABLE 3A25: IMPACT OF URHVP ON LIKELIHOOD OF CHWS PARTICIPATION IN SERVICE PROVISION AT HEALTH FACILITY

	(1) Participate in Community Health Programs	(2) Visit Community members	(3) Provide reports to Health Facility	(4) Attend Community health meetings & training	(5) Participate Community Health Programs	(6) Visit Community members	(7) Provide reports to Health Facility	(8) Attend Community health meetings & training
During URHVP (post1)	0.092*	0.057	0.018	0.044	0.082	-0.191**	0.044	0.000
	(0.039)	(0.037)	(0.031)	(0.033)	(0.057)	(0.048)	(0.061)	(0.046)
Program × Post1	**0.051**	**0.140**	**0.145**	**0.154***	**0.061**	**0.110**	**0.112**	**0.246****
	(0.087)	**(0.080)**	**(0.089)**	**(0.062)**	**(0.097)**	**(0.105)**	**(0.116)**	**(0.075)**
Withdrawal URHVP (post2)	0.069	0.103	-0.005	0.055	0.053	-0.392**	0.052	-0.027
	(0.044)	(0.055)	(0.044)	(0.043)	(0.068)	(0.054)	(0.082)	(0.069)
Program × Post2	**-0.089**	**-0.130**	**-0.036**	**-0.082**	**-0.107**	**-0.217**	**-0.104**	**0.055**
	(0.091)	**(0.085)**	**(0.103)**	**(0.070)**	**(0.136)**	**(0.109)**	**(0.163)**	**(0.138)**
Constant	0.596**	0.575**	0.543**	0.558**	0.588**	0.569**	0.545**	0.549**
	(0.027)	(0.026)	(0.023)	(0.021)	(0.012)	(0.012)	(0.012)	(0.010)
R-Wolf p-value (prog×post1)	[0.149]	[0.089]	0.149]	[0.089]	[0.347]	[0.347]	[0.347]	[0.020]
R-Wolf p-value (prog×post2)	[0.980]	[0.980]	[0.960]	[0.980]	[0.871]	[0.218]	[0.871]	[0.871]
R.sq.	0.041	0.070	0.044	0.059	0.214	0.228	0.151	0.196
Obs.	397.000	397.000	397.000	397.000	397.000	397.000	397.000	397.000
Facility Fixed Effects	Y	Y	Y	Y	Y	Y	Y	Y
District time trends	N	N	N	N	Y	Y	Y	Y
t-test (p-value)								
post1=program×post1	0.707	0.425	0.243	0.205	0.887	0.045	0.694	0.041
post2=program×post2	0.171	0.064	0.817	0.164	0.438	0.294	0.528	0.695
program×post1=program×post2	0.034	0.003	0.021	0.001	0.052	0.000	0.027	0.051

Notes: Outcomes from URHVP Health Facility Survey. Regressions in 5-8, control for district linear specific time trends. Participation in CHW service provision at health facility taking 1 for responses "Yes, all" and 0 for responses "Yes, some or No". Family-wise p-values (Romano-Wolf P-values), for the interaction terms (Program×Post1 and Program×Post2) are reported in square brackets (Clarke et al 2019). T-test (p-value) for statistical difference between coefficients for period and the interaction term for period with program. Standard errors are clustered at the district level with health facility fixed effects. Standard errors in parentheses. Significance levels at * $p < 0.05$, ** $p < 0.01$

TABLE 3A26: IMPACT OF URHVP ON LIKELIHOOD OF PROVISION OF QUALITY HEALTH SERVICES BY CHWS

Panel A	(1) 1 if CHW services are Very good	(2) 1 if CHW services are Very good
During URHVP (post1)	0.001	-0.051
	(0.034)	(0.037)
Program × Post1	**0.204***	**0.258****
	(0.075)	**(0.089)**
Withdrawal URHVP (post2)	0.036	-0.065
	(0.043)	(0.041)
Program × Post2	**0.027**	**0.130**
	(0.052)	**(0.082)**
Constant	0.192**	0.193**
	(0.019)	(0.015)
R.sq.	0.057	0.171
Obs.	397	397s
Facility Fixed Effects	Y	Y
District time trends	N	Y
t-test (p-value)		
post1=prog×post1	0.035	0.009
post2=prog×post2	0.922	0.123
prog×post1=prog×post2	0.036	0.141

Panel B	(1)	(2)
During URHVP (post)	-0.107	-0.027
	(0.105)	(0.374)
Program × Post	**0.536****	**0.211**
	(0.135)	**(0.193)**
Constant	-1.617**	-2.311**
	(0.527)	(0.627)
R.sq.	0.349	0.457
Obs.	292	292
Mother Fixed Effects	Y	Y
Controls	Y	Y
District time trends	N	Y
t-test (p-value)		
Post=Prog ×Post	0.004	0.642

Notes: Outcomes in Panel A from URHVP Health Facility Survey while outcomes in Panel B from URHVP Women Survey. In Panel A, standard errors are clustered at the district level with health facility fixed effects while in Panel B, standard errors are clustered at the district level with women fixed effects. Standard errors in parentheses. All regressions in Panel B control for woman's age at pregnancy, dummy for 1st born dummy, dummies for primary source of income for woman and spouse, dummies indicating possession of agricultural land, livestock, poultry and durables (TV, radio, motorcycle, bicycle, house). Regression in 2 controls for district linear specific time trends. Very good level of service provision equals 1 for responses "Very Good" and 0 for responses "Very Poor or Poor or Fair or Good". T-test (p-value) for statistical difference between coefficients for period and the interaction terms of period with program. Significance levels at * $p < 0.05$, ** $p < 0.01$

APPENDIX 3B: CONSTRUCTION OF ASSET INDEX USING PRINCIPAL COMPONENTS ANALYSIS

To proxy for economic status, following the step by step process of Fry K. et al. (2014), I use the CHW information on the number of different assets 'n' (a_{1j} to a_{nj}) owned by each CHW household 'j' in the pre-program period[52]. For each asset variable, I defined a dummy equaling 1 if the CHW owned the asset, and 0 if otherwise. I obtained the factor weight (f_1), which is the first principal component for each asset variable. I then standardized each variable by subtracting the mean (a_1) of that variable from each value (a_{1j}) and dividing it by the standard deviation (s_1). I then computed the asset index (A_{1j}) for each CHW by multiplying the standardized variables with the factor weight, and then adding all the values together. The asset index thus obtained is the sum of all the weighted standardized asset variables for each CHW, following Filmer and Pritchett (2001), as expressed in the equation below:

$$A_{1j} = f_{11} \times (a_{1j} - a_1)/(s_1) + \ldots + f_{1n} \times (a_{nj} - a_n)/(s_n)$$

I then classified and assigned CHWs to quintiles (1 to 5) based on their value on the index, where 1 is the poorest quintile and 5 the wealthiest quintile. I refer to the bottom 40% (quintile 1 and 2) as "poor," the next 40% (quintile 3 and 4) as "middle," and the top 20% (quintile 5) as "rich." For the purposes of this analysis, I consider the bottom 40% as poor taking 1, and 0 if otherwise (poorv), expecting that the effect of the program may be larger among poor CHWs than among those that were better-off. This approach is similar to Filmer and Pritchett (2001)

[52] Cattle, Goats/Sheep/Pigs, Poultry, TV, Radio, Car, Motorcycle, Bicycle, Mobile phone, Agricultural land, and House