Essays on social networks in developing countries

A thesis submitted to the University of Kent in fulfilment of the requirements for the degree of Doctor of Philosophy in Economics

Hyunhoi Koo

Faculty Advisor: Dr Bansi Malde and Dr Guy Tchuente

University of Kent

Canterbury, Kent, United Kingdom

April 2022

Declaration

I declare that the Chapter 4 is jointly authored with my supervisor, Dr Bansi Malde. This thesis contains no material that has been submitted previously, in whole or in part, for the award of any other academic degree.

Acknowledgements

Firstly, I want to thank my supervisor Dr. Bansi Malde for her support throughout the duration of my PhD studies. Her help in my study was invaluable and she offered me great research opportunities. I would like also to thank my supervisor Dr. Guy Tchuente who gave meaningful feedback regarding the implementation of econometric models and backgrounds in this thesis. Furthermore, I would like to thank all of my PhD colleagues and staffs at the School of Economics for their support. Lastly, I am grateful for the continuous encouragement from family and friends.

Abstract

Social networks can be vital in developing countries where institutions and markets are missing or incomplete. Social networks can supplement markets that are not suitable for people's economic needs by sharing resources, shaping beliefs and transmitting information. This thesis studies social networks in developing countries, which is important to find effective approaches to deal with the problems. The first chapter examines the peer effects of religious groups on learning one's own HIV status. This analysis is interesting because religion substantially affects the lives of people in sub-Saharan Africa and there could be peer effects within religious groups which drive behaviour including stigma. It is important to understand these in order to design policies to overcome stigma. I use exogenous variation from randomised incentives to learn test results in rural Malawi (Thornton, 2008) to address difficulties in disentangling peer effects from other effects. I find that the proportion of religious group learning their HIV results increases the probability of a group member's finding out their HIV test results. The second chapter studies peer effects on subjective expectations of HIV. Subjective expectations significantly influence health behaviours and play a crucial role in health outcomes by inducing take-up of preventive health behaviours and avoidance of risky behaviours. I use data from Delavande and Kohler (2009) which collected subjective expectation on one's own HIV status by easy visual methods in rural Malawi. The results show significant and positive peer effects on one's own subjective likelihood of HIV infection. The third chapter explores the effect of rainfall shock on within-village extended family networks. Within-village extended family networks, and the interactions within these networks, play an important role in risk and resource sharing. The influences of rainfall shocks – measured as absolute deviation from long-term average – on the structure of within-village extended family networks in rural Mexico are examined. In villages where reduction in incomes is implied due to rainfall shocks, extended family networks have smaller degree and size. However, these negative effects are countervailed by cash transfers. Possible explanations for these results are the decrease of females leaving their households for marriage in control group and reduced work migration in treatment group with cash transfers. The findings in this thesis emphasise the importance of social networks in developing countries in terms of behaviour, belief and risk-sharing.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Social networks can be vital in developing countries where institutions and markets are missing or incomplete (Breza et al., 2019). Social networks can supplement markets that are not suitable for people's economic needs by sharing resources, shaping beliefs and transmitting information. In developing countries, where the provision of essential health products and formal institutions is sub-optimal, social networks can be important for adopting health products or reacting to economic risks. Therefore, the economics of networks has emerged as one of the powerful approaches to addressing issues threatening economic development in developing countries, such as public health problems and weather events (Kohler et al., 2007; Angelucci et al., 2009; Godlonton and Thornton, 2012; Kinnan et al., 2012; Corno et al. 2020). Social networks can help or hinder the effectiveness of different policy interventions. For example, health interventions with information provision via social networks can be an effective method for HIV prevention when friends and family members are the sole sources of information in developing regions (Dupas, 2011; Hoffman, 2017). However, social stigma prevalent among social networks is an obstacle to scaling up HIV testing in sub-Saharan Africa (Novignon et al., 2014). Studying networks in developing countries is therefore important to find effective approaches to deal with the problems. This thesis studies the incentives to obtain HIV test results, the formation of health beliefs and extended family network characteristics to study the effect and characteristics of social networks in terms of economic development and peer effects.

The first chapter examines the peer effects of religious groups on learning about HIV status. Religion substantially affects the lives of people in sub-Saharan Africa through high participation in religious activities and trust in religious leaders. Thus, religious groups play a significant role in this region. Furthermore, religious lessons related to health behaviours, often prescriptive, could be shared through religious group members and generate homogenous health behaviours. For instance, the Catholic Church has been opposed to condom use, notwithstanding their importance for offering protection against the spread of HIV. Stigma has been considered an obstacle in promoting HIV tests in sub-Saharan Africa. Religious leaders could have reinforced the stigma against HIV and disincentivised HIV testing by spreading messages within religious groups. Therefore, the first chapter examines the peer effects of religious groups on decisions to learn HIV test results, which might be a possible approach to overcoming stigma. In the experiment conducted by Thornton (2008) in rural Malawi, monetary incentives for an individual to learn about their test results are randomly assigned. This study used exogenous variation from these randomised incentives to address difficulties in disentangling peer effects from other effects, such as similar characteristics of group members. The results reveal a one-percentage-point increase in the proportion of a religious group learning HIV results increases the probability of finding out about HIV test results by 0.23 percentage points. This finding suggests the possibility of exploiting religious networks for scaling up HIV testing in sub-Saharan Africa.

The second chapter studies peer effects on subjective expectations of HIV. Subjective expectations significantly influence health behaviours and play a crucial role in health outcomes by inducing take-up of preventive health behaviours and avoidance of risky behaviours. For example, downward (upward) revisions in subjective beliefs about HIV infection increased (decreased) risky behaviour (Gerrard et al., 1996; Paula et al., 2014). In developing countries, where people lack health services and education, social networks are potentially the critical sources of information affecting health beliefs and expectations. However, an individual's subjective expectations of their HIV infection might be formed through incorrect health knowledge based on social norms or traditions. Thus, estimating peer effects on the subjective likelihood of HIV infection with relevant subjective expectation data could be important. I use data from Delavande and Kohler (2009), using visual tools and easy methods to collect subjective expectations precisely in rural Malawi. Estimations using this data exploit the variation of randomised monetary incentives offered from HIV testing to identify peer effects apart from other unobservable factors. The results show significant and positive peer effects on individuals' subjective likelihood of current and future HIV infection. Furthermore, peer effects are larger for those who learned of their HIV test results. These findings emphasise the importance of social networks for correcting the overestimation of HIV risk and the need to promote HIV testing.

Finally, the third chapter explores within-village extended family networks and rainfall shock. Studies suggested strategic network formation through marriage or migration for risk-sharing as a coping mechanism in developing countries for risk (Rosenzweig and Stark, 1989; Henry et al., 2004; Puente et al., 2015). Within-village extended family networks, and the structure of interactions within these networks, play an important role in risk and resource sharing. However, not many studies explored how extended family networks are affected by income shocks. This chapter studies the influence of rainfall shocks – measured as absolute deviation from long-term average rainfall – on the structure of within-village extended family networks in rural Mexico. In villages where a reduction in agricultural incomes is implied due to rainfall shocks, extended family networks are likely to have smaller degree and size. However, within treatment villages with cash transfers, these effects of rainfall shocks on degree and size are countervailed. Possible explanations for these effects are decreased split households implied by a reduction in females leaving their household for marriage in the control group and reduced work migration in the treatment group after rainfall shocks, suggesting that cash transfers influence the structure of these networks.

Chapter 2

Peer effects of religious groups upon learning HIV status

2.1 Introduction

HIV testing is a vital component of any HIV control policy, yet widespread social stigma is thought to constrain voluntary testing by individuals (Novignon et al., 2014). Thus, it is critical to understand the drivers of such stigma to identify effective solutions to scale up HIV testing. Religion significantly affects the lives of people in sub-Saharan Africa (Green, 2003; Iyer, 2010). However, whether religion can positively or negatively affect HIV prevention is debatable. For instance, many religious leaders opposed the use of condoms since it is perceived that their use may encourage unethical sexual activity, despite the HIV pandemic (Benagiano et al., 2011). Furthermore, the ties within a religious community are likely to be stronger than those in the neighbourhood, and religious groups are homogenous in terms of health-related beliefs or behaviours (Maselko et al., 2011). Therefore, peer effects could be stronger within religious groups which drive behaviour including stigma than influences in spatial neighbourhood. Furthermore, targeting religious groups could be an effective approach to overcome stigma and affect behaviour for HIV prevention through peer effects in sub-Saharan Africa. In this chapter, the role of religious groups in the decision to learn about HIV test results is examined.

The peer effects of religious groups are estimated using data from The Malawi Longitudinal Study of Families and Health (MLSFH) conducted in rural Malawi. Malawi is one of the least developed countries with a high HIV prevalence. The MLSFH collected long-standing and exhaustive data in data collection rounds in 1998, 2001, 2004, 2006, 2008, 2010 and 2012 for up to 4,000 individuals selected to represent the rural population of Malawi (Kohler et al., 2014). Randomised monetary incentives in 2004 to learn HIV test results provide the source of variation to identify peer effects. Furthermore, information on religious leaders is available to create religious group data. I estimate peer effects within religious groups on the decision to learn one's own HIV test results.

Disentangling peer effects from the behaviour in networks is considered difficult because the effects can also be derived from the influence of exogenous peer effects and the similar characteristics of group members. I use randomised monetary incentives to encourage obtaining test results as instrument variables to distinguish the role of religious group members on individuals' learning test results. This strategy is used by Godlonton and Thornton (2012) to examine peer effects in geographical networks. I use a similar identification strategy to examine the effect in religious groups, which may be more significant in affecting network partners. This methodology will also allow for assessing whether the drivers of peer effects are within villages or within religious groups since religious groups span multiple villages.

I find significant and positive peer effects of religious group members on one's learning HIV test results. A one-percentage-point increase in the proportion of religious groups learning about test results increases an individual's probability of obtaining test results by 0.23 percentage points, which is larger than the effects in the spatial network (0.12) measured by Godlonton and Thornton (2012). This finding implies that the role of religious networks is possibly larger than that of geographical neighbours in HIV prevention. When peer effects are re-estimated within groups separated by both religious groups and villages, effects are only significant for those in the same religious groups and villages. This finding implies that the effects are derived from interactions within religious groups rather than whether group members hold similar views. More religious group members in the same village would mean that there are more people to enforce a religious message or norm. Thus, influence from the combination of spatial networks and religious groups can be more powerful in changing HIV-related behaviour. These findings suggest that targeting religious groups can be an efficient approach to HIV prevention in sub-Saharan Africa. Providing incentives for religious group members could increase the impact through peer effects. The approach with religious groups shows the possibility of enhancing HIV testing by overcoming the stigma prevalent in religious networks. Furthermore, peer effects infer that incentives can only be provided to a subset of the group, instead of all group members, which is efficient where there is a limited budget for intervention.

2.2 Literature

2.2.1 HIV testing

HIV testing is an integral part of any treatment protocol since it is necessary for identifying those with HIV in order to be able to treat it. Significant literature has studied what gives rise to HIV testing, the effect of HIV testing, and how to scale up such testing in developing countries. Ngatia (2011) claimed that failure to diagnose HIV would cause people to be unaware of the risk of transmission and have consequences for society. Thus, timely HIV testing is indispensable for public health and is strongly recommended as an international public health policy. Researchers indicate that the stigma associated with HIV engenders fear in being identified with HIV, thereby stopping people from getting tested. This fear also hampers those who learn about their HIV test results from recommending testing to others because such testing might make others assume that those being tested have already been diagnosed as HIV positive (Novignon et al., 2014). These studies emphasise how stigma can be a strong obstacle since it prevents individuals from getting tested regardless of their HIV status. Studies have shown the possibility of an individual's HIV test result changing their sexual behaviour. With the data on people randomly assigned for HIV testing in Kenya and Tanzania, Gong (2014) showed evidence that those HIV-negative who had believed they were HIV-positive reduced their risky sexual behaviour. Thornton (2008) asserted that learning one's HIV-positive status significantly increases condom purchases. These studies indicated a possible effect of getting HIV tests on sexual behaviour, leading to possible HIV prevention.

Therefore, the goal for public health might be to overcome social barriers and stigma, leading more people to get HIV tested and learn about their HIV status. Thornton (2008) found cash incentives an effective means to scale up HIV testing even within a stigmatised society. By offering monetary incentives for HIV tests in three regions in Malawi, the experiment showed that willingness to learn about HIV test results was highly responsive to the incentives. Also, the costs of obtaining HIV test results, such as travel distance to the voluntary counselling and testing (VCT) centre, played a significant role in learning about HIV status. This finding is powerful in that stigma, one of the strong barriers to learning about HIV status, can be solved by providing small amounts of incentives.

2.2.2 Networks and HIV prevention

Concerning HIV, interactions in social networks and peer effects can play a significant role for people in sub-Saharan Africa. Social networks can help those in sub-Saharan Africa to learn about the causes and consequences of HIV and relevant approaches to prevent contracting it. Kohler et al. (2007) collected longitudinal data on HIV and social communication related to HIV in rural Kenya and Malawi and showed that social networks have causal effects on individuals' HIV risk perceptions. This empirical analysis suggests that the impact of social networks is important to understand the behavioural change in response to HIV. Furthermore, the effects of the number of network partners with different risk perceptions on respondents' personal risk perceptions were significant.

A possible approach to evaluate the impact of social networks on HIV prevention is to examine peer effects on behaviour related to HIV prevention. Godlonton and Thornton (2012) presented a good example of this approach. The study uses the data from an experiment that allocated financial incentives to individuals in rural Malawi to learn about their HIV test results. The main specification estimates the impact of geographical neighbours learning their HIV test results on individuals learning their HIV test results. The main finding of this study is that a ten-percentage-point increase in the portion of neighbours within 500 metres of learning HIV test results significantly raises an individual's probability of learning results by 1.1 percentage points. This finding is interesting in that the effect of geographical networks on behaviour, which is beneficial for HIV prevention, is evaluated in the rural settings of a developing country with a high HIV prevalence rate. It could be more relevant to examine the effect within religious networks due to the importance and homogenous beliefs within religious groups in sub-Saharan Africa.

Although exploring social networks is essential in studying HIV prevention in sub-Saharan Africa, it is challenging to identify the effects of networks. The reflection problem makes it difficult to distinguish the exogenous effects of peer characteristics from the endogenous effects of peer outcomes (Manski, 1993). Unobserved correlated characteristics are another factor that makes peer effects harder to identify. Furthermore, the effects might be driven by the fact that individuals have been self-selected into the network, referred to as endogenous network formation. Researchers have sought to address these identification challenges by adopting strategies such as instrumental variables for network effects or randomised experiments. Godlonton and Thornton (2012) took advantage of a randomised experiment to identify endogenous peer effects. Exogenous monetary incentives to encourage learning about HIV test results are used as instrumental variables to disentangle the effect of neighbours learning their HIV test results from other effects. Respondents were randomly and independently offered different amounts of money, which could motivate them to learn about their HIV test results. Exogenous monetary incentives offered to respondents significantly impacted their behaviour surrounding whether to learn about their HIV test results. In addition, an individual's decision in this regard is only affected by their monetary incentives but not influenced by the incentives assigned to others. Thus, monetary incentives satisfy the exclusion restriction since incentives afforded to others only affect one's decision to obtain test results through others' learning about their test results.

2.2.3 Religion and religious network

Researchers highlighted the role of religious organisations and leaders on the health behaviour and HIV-related consequences of religious people because religious messages can offer advice for preventing HIV and lead to safe and healthy behaviours (Garner, 2000). Religious organisations are deeply concerned about the socialisation of people. Thus, religious spaces are not only venues for teaching values and practices but also an environment for enhancing the sexual health of people (Paiva et al., 2010). Moreover, religious leaders are trusted individuals in society, so their messages are effective, and their actions can serve as an example (Oluduro, 2010). Adams and Trinitapoli (2009) claimed that relationships in a congregation might contribute substantially to how these relationships affect HIV prevention because denominations are normally characterised by formal hierarchical structures. Therefore, the discussion or intervention can be restrained and permitted in the congregation. The project collected data on the structure and the content of religious institutions in rural areas within three districts in Malawi, focusing on their efforts to address HIV in their organisations. Through this qualitative study, Adams and Trinitapoli (2009) assessed that religious leaders in these regions are actively engaging in HIV prevention in various ways and affecting how the people in their organisations are confronting HIV.

Religions can facilitate HIV prevention by sending prescriptive messages about a relevant attitude towards HIV and diffusing behavioural change. On the other hand, the strong influence of religious leaders can trigger stigma by labelling people with HIV as 'immoral' or 'promiscuous'. Due to stronger ties and homogenous thoughts in religious communities, religion can have both beneficial and deleterious effects on HIV prevention (Maselko et al., 2011). Furthermore, religious lessons can be an obstacle to the diffusion of safe sexual behaviour, such as condom use. In particular, the Catholic Church has been opposed to the 'artificial contraception', including condom use, and this position has not been changed even with the advent of the HIV pandemic (Benagiano et al., 2011). However, Trintapoli (2009) found that people guided by religious leaders who accept condom use are more likely to respond that they use condoms. In another study, Trintapoli (2011) examined the data from the Malawi Religion Project (MRP) and found that Catholic Church leaders were not more likely than Muslim group leaders to prohibit condom use among their members.

Some researchers paid attention to what religious networks can do for members in developing countries. In this regard, Debnam et al. (2012) asserted that social support could particularly benefit those in religious networks since those who attend religious services are more likely to be in large social networks and engage in support. Religious networks play a substantial role even in non-religious activities in developing countries, such as in health and education (Iyer, 2010). Religious networks are unique in that religious communities formally involve a gathering of individuals with similar beliefs and backgrounds. Furthermore, religious network partners can strengthen people's ties through longstanding relationships and shared beliefs. Religious networks can accept homogeneous thoughts and behaviours in terms of health and exert a stronger impact on health (Maselko et al., 2011).

Despite the important role of religious networks on healthy lives in developing countries, not many researchers have measured the impact of religious networks on health outcomes. Most studies focused on religious leaders. However, some studies pointed out that religious leaders are not necessarily the only element on which focus should be placed regarding the impact of religion on HIV prevention. Adams (2007) argued for the need to concentrate on the intra-organisational relationships among the members of congregations rather than on the hierarchical structure of those organisations when examining HIV prevention strategies.

Furthermore, Trinitapoli (2007) investigated the survey data on unmarried adolescents in rural Malawi and found that frequent messages about HIV from religious leaders do not significantly affect adolescent abstinence and the sexual behaviour of the members of religious organisations.

2.3 Background and data

2.3.1 Background and survey design

I use the data from the Malawi Longitudinal Study of Families and Health (MLSFH) conducted in three regions of Malawi to examine the role of religious networks in the decision to get HIV test results. The MLSFH is a longitudinal cohort study based in three rural districts in Malawi (Rumphi in the north, Mchinji in the centre, and Balaka in the south). A large amount of data is collected through seven survey rounds from 1998 to 2012 to study the health, society, economics and demographics of the population in rural Malawi. The topics cover such topics as family dynamics, human capital, social networks, sexual behaviours and subjective expectations, consumption (Kohler et al., 2007).

Figure 2.1: MLSFH locations (Kohler et al., 2014)

Malawi is a land-locked country in southern Africa. It is considered a least developed country with a large rural population. Approximately 90% of people in Malawi live on a daily income below US$ 2 per day (Emina et al., 2013). UNDP (2019) reports that Malawi's Human Development Index ranking is 172 among 189 countries. According to the World Health Organization (WHO), the HIV prevalence rate among adults 15–49 years old in Malawi was 9.2% in 2018, which is above the average global rate (0.8%) and higher than the average prevalence rate in Africa (3.9%). Therefore, studies in rural Malawi can represent sub-Saharan African conditions with a high HIV prevalence, unfavourable health environment and facilities and low standards of living (Kohler et al., 2007).

The sample for the MLSFH was selected to represent the rural Malawi population. Individuals aged 15–49 were randomly selected from approximately 120 villages in three rural regions and were surveyed in the first survey round in 1998. Approximately 25% of all households in each village participated. During data collection in 2004, an additional sample of young adults aged between 15 and 24 was added to the sample. This sample addition was necessary to compensate for the ageing of the initial sample and to supplement adolescents. The MLSFH provides longstanding data on health, economic and demographic conditions in sub-Saharan Africa from 1998 to 2012 for up to 4,000 cohorts.

Between May and August of 2004, respondents were offered free tests for HIV and other sexually transmitted infections (STIs). Before tests for HIV were offered, respondents were provided with counselling about HIV prevention strategies. The survey was conducted, and various data, including the demographics of respondents, were collected. If respondents accepted HIV testing and got tested, the results of the tests were available 4–6 weeks after the test for those tested. There was a possibility that respondents accepted HIV tests but did not obtain their test results. Thus, respondents were offered randomly assigned monetary incentives to get their test results. Monetary incentives ranged from zero to 300 Kwacha. The monetary incentives were randomly assigned with variations of 50 Kwacha (e.g. 50, 100 and 150). Kwacha is the currency of Malawi and the nominal exchange rate was 108.9

Malawi kwacha per US dollar in 2004 (Pauw et al., 2013). Therefore, monetary incentives varied from approximately US$ 0 to US$ 3. Respondents collected their test results from the VCT centres two months later. The distances to VCT centres were randomly located. Additional surveys were performed every two years until 2012, so respondents were subjected to biennial re-interviews. Detailed information on religion and religious leaders was collected in the 2008 survey.

I use mainly the data collected before respondents got HIV tested in 2004. Furthermore, I added data from the 2008 survey round, including religion and religious leaders, to generate religious network data. Religious network data was generated by forming religious groups comprising respondents. Religious groups are defined as groups of individuals under the same religious leaders in the same religion and region. It is necessary to limit religious groups in the same religion because a couple of religious leaders have the same names but belong to different religions. Religious groups under the same religious leaders are likely to be the closest group available from the data because large churches or mosques might exist, and not all individuals within the same religious organisations are likely to be acquainted. Those in the same religious groups are highly likely to conduct religious activities together and communicate regularly. The information on villages was also available. However, some respondents reported that they attend churches or mosques in villages in which they do not live. Thus, religious groups were formed regardless of the villages where group members lived. Religious and geographical networks do not overlap in this analysis.

2.3.2 Sample selection

The sample used for the analysis in this chapter consists of individuals who accepted an offer to get HIV tested in 2004 and interviewed in 2008 with no missing data on monetary incentives for HIV testing and religion. Next, the main sample included those who did not change religion between 2004 and 2008 and, in 2008, were in religious networks under the same religious leaders. This approach is taken because the data on religious leaders are taken from 2008 data.

As a part of the MLSFH survey in the 2004 round, 2,265 respondents, randomly selected from rural regions in Malawi, were offered HIV tests. Among the 2,265 respondents offered testing in three districts in Malawi, 92% accepted getting HIV tested. Among those without missing data on monetary incentives to encourage receiving their test results, 1,997 individuals had no missing data for religion in both 2004 and 2008. Among respondents whose religion data was available, 1,805 respondents who did not change religion between two survey rounds were selected. Ten respondents who responded that they do not have a religion and 89 respondents who reported they do not have religious leaders were dropped from the sample. Finally, 1,045 respondents were selected as the main sample because of their membership in religious groups under the same religious leaders. Furthermore, 661 respondents were dropped because they were in singleton groups without other group members. Therefore, the limitation of the data might be that the data of respondents who changed religion are excluded. However, only 8% (197 individuals) of respondents who are offered tests (2,265 samples) changed religion in this data.

Table 2.1: Sample selection

Selection process	N
Offered HIV test	2,265
Accepted HIV test	2,075
Incentive for HIV test is not missing	2,062
Religion is not missing	1,997
Religion did not change between 2004 & 2008	1,805
Have religion	1,795
Have religious leaders	1,706
Main sample (in religious groups)	1,045

Table 2.2 presents the results of pairwise t-tests to compare characteristics (gender, age, marital status, education, land ownership and HIV status) between respondents who were dropped and selected in each sample selection process. These characteristics present basic properties of samples and are included in peer effects estimation as control variables. Most selection processes do not change characteristics significantly. Those who did not change

religion are more likely to be older and married compared to those who changed religion between 2004 and 2008. This tendency might be related to adolescents being added in the sample for the survey rounds after 2004. Furthermore, it might also be because individuals tend to change religion when they get married. However, age and marital status are not substantially different between dropped respondents who refused to get tested and the main sample. Gender and HIV status are significantly different between dropped and selected samples on whether an individual has a religion. However, the number of dropped samples is very small (ten out of 1805). Except for sample selection steps with small amounts of sample drops, there are almost no significant differences in characteristics between dropped samples and selected samples.

Table 2.2: P-values from pairwise t-tests for sample selection

	Accepted test			Incentive for test not missing			Religion in 2004 not missing			Religion in 2008 not missing		
	Drop	Sel.	p	Drop	Sel.	p	Drop	Sel.	p	Drop	Sel.	p
Male	0.43	0.44	0.82	0.38	0.44	0.67	0.57	0.44	0.37	0.35	0.44	0.13
Age	34.6	34.1	0.72	33.1	34.1	0.75	33.4	34.1	0.83	36.9	34.0	0.21
Married	0.78	0.75	0.28	0.69	0.75	0.59	0.71	0.75	0.77	0.71	0.75	0.44
Education	3.50	3.89	0.20	2.43	3.90	0.11	3.67	3.90	0.81	3.39	3.91	0.15
Owns land	0.76	0.76	0.93	0.56	0.76	0.17	0.86	0.76	0.46	0.71	0.76	0.41
HIV+							0.08	0.04	0.64	0.06	0.04	0.56
N	190	2075		13	2062		14	2048		51	1997	
	Religion not change			Have religion			Have leaders			In singleton religious groups		
Male	0.43	0.44	0.78	0.90	0.44	0.00	0.46	0.44	0.79	0.46	0.43	0.27
Age	29.4	34.5	0.00	30.7	34.6	0.41	32.1	34.7	0.28	34.6	34.7	0.83
Married	0.57	0.77	0.00	0.60	0.77	0.27	0.70	0.78	0.20	0.77	0.78	0.44
Education	4.28	3.87	0.10	3.50	3.87	0.68	3.96	3.86	0.79	3.61	4.02	0.03
Owns land	0.71	0.77	0.12	0.86	0.77	0.50	0.74	0.77	0.60	0.78	0.76	0.42
HIV+	0.03	0.04	0.19	0.00	0.04	0.00	0.02	0.04	0.22	0.05	0.04	0.71
N	192	1805		10	1795		89	1706		661	1045	

The initial selection process of respondents offered HIV testing was random, and there were no incentives offered to participate in the survey or test. Therefore, adverse selection, such as a possibility for individuals whose behaviour is risky to participate in the first place, or self-selection into the process due to monetary incentives, is not a serious concern in this sample selection. Furthermore, measurement error may occur because all individuals in religious groups are not included in the main sample. However, this may not be the issue for the instrumental variable regression since instruments are randomised monetary incentives. This experiment, involving randomly assigned incentives, would also randomise the effect of individuals who do not exist in the main sample.

Selection bias can be an issue since the main sample is selected from those offered HIV tests in 2004, with religion and religious leaders measured in 2008. The possibility of a selection problem can be critical if the choice of religion is affected by the test results and the decision to learn test results. In order to test the possible endogenous group formation, I examine the effect of behaviour and results of HIV tests in 2004 on the responses about religion in 2008. The results in Appendix 2A do not find significant effects on the choice of religion.

The choice of religious leader might be also related with test results and incentives to learn test results. It would be ideal to check the exogeneity of group formation caused by the choice of leader. However, the data on religious leader is not available for survey in 2004. Examining the relationship between test results (and incentives) and the choice of religion was possible approach to test endogenous network formation and sample selection due to data limitation. Using the choice of religion could be relevant since there are not much variation of religious groups within religion in local villages. In the data for the analysis, there are two religious groups, on average, in each religion within local village. Thus, it might be difficult to switch religious leader without switching the religion.

2.3.3 Data

Table 2.3 presents the descriptive statistics of the main sample comprising 1,045 individuals. Forty-three per cent of the main sample are male. Respondents selected for the main sample are, on average, 35 years old with, on average, four years of education. Seventy-eight per cent of the main sample is married, and 76% owns land. Religions are classified as Catholic, Church of Central Africa Presbyterian (CCAP), other Christian (e.g. Baptist, Anglican, Pentecostal, 7th Day Adventist) and Muslim. Seventeen per cent of the main sample is Catholic, and 19% is CCAP. The categories titled 'other Christian' and 'Muslim' have the highest portion (30%) of respondents. Examples of 'other religions', which represent 5% of the main sample, are New Apostolic, African Continent and Bible Believers.

Table 2.3: Descriptive statistics

		N	Mean	SD
Demographics	Male	1045	0.43	0.50
	Age	1045	34.73	13.32
	Married	1045	0.78	0.41
	Years of education	1045	4.02	2.92
	Owns land	1045	0.76	0.43
Religion	Catholic	1045	0.17	0.37
	CCAP	1045	0.19	0.39
	Other Christian	1045	0.30	0.46
	Muslim	1045	0.30	0.46
	Other	1045	0.05	0.21
HIV test	Respondent learn results	1045	0.73	0.44
	Amount of incentive[1]	1045	109.53	97.39
	Received a non-zero incentive	1045	0.77	0.42
	HIV positive	1045	0.04	0.20
Group average	Respondent got results	237	0.75	0.27
	Amount of incentive	237	108.90	58.92
	Number of members	237	4.41	4.37

[1] Kwacha ≈ 0.01 USD

[2] The main sample consists of individuals who accepted an offer to get tested without missing data. They are members of 237 religious groups.

Among 1,045 individuals who accepted HIV testing and were tested, 73% obtained their test results. Individuals who accepted HIV tests were offered monetary incentives, randomly assigned between zero and 300 Kwacha, to get their test results. The maximum incentive amount, 300 Kwacha, is approximately US$ 3 (Pauw et al., 2013). This amount is, on average, comparable to two days' wages for rural Malawians (Fedor et al., 2015). The value of incentive offered to those in the main sample is 110 Kwacha (approximately US$ 1.1) on average, and 77% received a non-zero incentive. As a result of HIV tests, 4% of the main sample tested HIV positive. This finding is lower than the national HIV prevalence rate (12% among adults 15-49 years old in 2004) (UNAIDS, 2019). This difference is normal for prevalence rates from a population-based study since national prevalence rates are typically collected from clinics (Godlonton and Thornton, 2012).

The respondents in the main sample come from 237 religious groups. Religious groups consist of four members on average, and the number of members varies from two to 28. The average amount of monetary incentives to learn test results for each religious group is 109 Kwacha (around US$ 1) on average. However, the average value of incentives afforded to religious groups varies from 0 to 300 Kwacha. There are only six groups where all members are not assigned monetary incentives, and one group with all (two) members given 300 Kwacha.

2.4 Estimation

2.4.1 Identification strategy

An indispensable procedure in seeking to measure the impact of religious networks on the decision to learn HIV test results is to develop a strategy to address difficulties in evaluating the effects of social networks. It is not very easy to distinguish the role of networks since an individual's behaviour can be driven by the influence of exogenous peer effects, the behaviour of individuals in the network and the fact that individuals in the same network behave similarly due to their similar characteristics (Manski, 1993). Similar characteristics in network groups can be formed through self-selection into networks. In this estimation, the respondent's decision on HIV testing can be driven not only by the decision of other group members but also similar characteristics and unobserved factors common across group members.

In the experimental design of the 2004 MLSFH survey round, individuals tested for HIV were offered a randomly assigned amount of monetary incentive to obtain their test results. Therefore, the strategy of using instrument variables is feasible due to the randomised monetary incentives. Monetary incentives to encourage respondents to obtain test results are exogenous because the amount is randomly assigned to them. An individual's behaviour in learning about their HIV test result is expected to be affected by the amount of monetary incentive assigned. Randomly assigned monetary incentives to obtain test results satisfy instrument relevance. In addition, the individual's decision to get their test results is not directly affected by the incentives of other group members. Incentivising religious group peers affects the decision of group member to obtain test result only through the effect of these incentives on religious group peers learning their test results, thereby validating the exclusion restriction of the instrumental variable. Randomised incentives is a good instrument, provided religious group formation is exogenous and does not depend on these incentives. The finding that the choice of religion is not affected by test results and incentives to learn test results discussed in 2.3.2 Sample selection provides a piece of evidence to relieve the concern on endogenous group formation.

2.4.2 Model

I regress an individual religious group member obtaining their HIV test result on their fellow religious group members learning of their respective HIV results as follows:

$$Y_{ij} = \alpha + \beta N_{ij} + \gamma X_{ij} + \alpha_j + e_{ij} \tag{2.1}$$

Y_{ij} is an indicator of whether individual i in a religious group j learned of their test result. N_{ij} is the proportion of members in an individual's i's religious group j who learned their results, X_{ij} is a vector of controls (controls are explained in the next paragraph) and α_j is a religious-group fixed effect. Leave-out means of X variables are included as controls. Leave-out means indicate the group average calculated without the individual's respective value. The main coefficient, β, shows the effect of the proportion of members in an individual's religious group who obtained their HIV test results on whether the individual learned their personal HIV test results.

The vector of controls includes basic demographic characteristics, including age, age-squared, gender, years of education, a dummy of whether a respondent owns land, marital status, religion and the number of members in the religious group. Control variables also include variables related to monetary incentives to learn test results, such as a dummy if a respondent is offered non-zero amounts of incentives. The distance to testing centres is important as these distances constitute costs in obtaining test results. Therefore, for dummies where the distance to VCTs where respondents obtain their test results is over 1.5km, district fixed effects and the average distance to simulated locations in the VCT zone are included in the vector of controls.[1]

Using the proportion of the group as an explanatory variable to estimate peer effects reflects how an individual tends to conform to their group members. Many researchers who studied peer effects apply the model, using the percentage of a certain group. For example, Godlonton and Thornton (2012) estimated the effect of geographical neighbours learning HIV test results on an individual obtaining their personal test results with the proportion of tested neighbours who learned their test results as an explanatory variable. In another study, Godlonton and Thornton (2013) examined the effect of neighbours getting tested on individual beliefs and behaviour by regressing dependent variables on the proportion of neighbours in a village who learned their HIV test results. The model is an example of a linear-in-means model, estimating the network effects of the portion of religious group members instead of the absolute number of religious group members who obtained test results. This type of model assumes that diffusion in networks occurs through conforming or imitating the group average instead of contacting individuals (Fafchamps, 2015).

[1] As calculated by Thornton (2008), the average distances to simulated locations in a VCT are included as a control for the distance to the VCT and the location of respondents in each VCT zone since the location of the VCT is chosen randomly in each zone. This control variable was included in several studies using the same 2004 MLSFH data related to HIV testing (Thornton, 2008b; Godlonton and Thornton, 2012).

2.4.3 First stage

I use randomly assigned monetary incentives to encourage respondents to obtain HIV test results as instrumental variables. The first stage of the regression is as follows:

$$N_{ij} = \alpha + \theta_1 IC50_{ij} + \theta_2 IC100_{ij} + \theta_3 IC200_{ij} + IC300_{ij} + \delta X_{ij} + e_{ij} \tag{2.2}$$

where instruments are the proportion of religious group members with the quantum of incentives to encourage them to obtain test results separated by various cut-off amounts denoted in Kwacha ($\approx$ 0.01 USD). $IC50_{ij}$ is the proportion of religious group members who are offered incentives from ten to 50 Kwacha and $IC100_{ij}$ is the proportion of religious group members who are offered incentives between 60 to 100 Kwacha. $IC200_{ij}$ and $IC300_{ij}$ are the proportion of religious group members with incentives from 110 to 200 Kwacha and from 210 to 300 Kwacha, respectively. The minimum value of incentives assigned is 10 Kwacha and the maximum value assigned is 300. I instrument the portion of religious group members who learned their HIV test result, N_{ij} in Equation 2.1.

These cut-off amounts are set to consider both a similar proportion of respondents and similar amounts of incentives. According to Table 2.4, the portion of respondents is distributed from 18% to 25% by the value of incentives divided by the cut-off from instruments. The portion of religious group members who are assigned monetary incentives from ten to 100 Kwacha is divided into two variables because the portion of the group (37%) was substantially higher than the others. This set of splines is used as the set of instrumental variables to allow flexibility in the relationship between instruments and variables, for example, non-linear relationships. A couple of other studies used this approach to allow flexible relationships between instruments and explanatory variables in analyses using the same dataset with randomised monetary incentives as instrument variables (Godlonton and Thornton, 2012)

Table 2.5 presents the result of first stage regression, for which estimates show that the

Table 2.4: Portion of individuals with incentives to learn test results

	Mean	SD
Non-zero	0.77	0.26
Between 10 and 50	0.19	0.23
Between 60 and 100	0.18	0.22
Between 110 and 200	0.25	0.25
Between 210 and 300	0.15	0.22

proportions of members with various values of incentives offered to religious group members significantly affect the portion of group members who obtain HIV test results. The F-statistics is 26.04. This result confirms the instrument relevance and validity of monetary incentives to obtain HIV test results.

Table 2.5: First stage for peer effects on learning test results

	Portion of members who learned results
Proportion of members with incentive 10-50	0.43***
	(0.07)
Proportion of members with incentive 60-100	0.58***
	(0.06)
Proportion of members with incentive 110-200	0.61***
	(0.07)
Proportion of members with incentive 210-300	0.54***
	(0.07)
N	1045
F	26.04

*, **, *** indicate significance at the 10, 5 and 1%. Covariates: age, age-squared, gender, years of education, land ownership, marital status, if offered non-zero incentive, the amount of incentive, district, region, whether the distance to the result centre (VCT) was over 1.5km, the number of group members, and average distance to simulated locations in VCT zone.

2.4.4 Religious network effects

Table 2.6 presents the ordinary least squares (OLS) and instrumental variables (IV) estimates for the peer effect of religious network partners. Column (1) shows the OLS estimates and Column (2) presents the IV estimates. The main coefficients of the proportion of the group to obtain test results are all positive and significant for both regressions. The IV estimates of peer effects imply that a one-percentage-point increase in the proportion of an individual's religious group learning HIV results increases the individual's probability of obtaining HIV test results significantly by 0.23 percentage points. The average number of group members is around four from the descriptive statistics (Table 2.3), implying that one group member obtaining a test result in a religious group on average increases another member's probability of learning their test results by 5.8 percentage points. The results suggest positive and significant peer effects from religious group members on an individual's decision to obtain HIV test results. The effects of incentives are also positive and significant. OLS estimates that are larger than IV estimates imply that the peer effect is overestimated without considering correlated effects and endogenous network formation. The large difference between OLS estimates and IV estimates demonstrates the need to use a relevant identification strategy to identify peer effects, accounting for all the possible sources of endogeneity.

Table 2.6: Peer effects in religious groups on learning test results

	Learn test results	
	OLS	IV
	(1)	(2)
Portion of members who learned results	0.82***	0.23**
	(0.05)	(0.11)
Non-zero incentives	0.26***	0.32***
	(0.04)	(0.04)
Amounts of incentives	0.00***	0.00***
	(0.00)	(0.00)
N	1045	1045
F	93.89	21.58

SE clustered at the village level. Covariates: age, age-squared, gender, years of education, whether has land, marital status, if offered non-zero incentive, the amount of incentive, whether the distance to the result centre (VCT) was over 1.5km, district, religion, the number of group members, and average distance to simulated locations in VCT zone.
*, **, *** indicate significance at the 10, 5 and 1%.

The estimated effects are larger in magnitude than the effects of geographical neighbours with the same dataset, as measured by Godlonton and Thornton (2012). Godlonton and

Thornton (2012) estimated the coefficient of 0.12 from the portion of neighbours within 0–0.5km who obtained test results. This estimate implies that the role of religious network partners is possibly larger than that of geographical neighbours in HIV prevention. The reason might be that religious networks have stronger ties than geographical neighbours since religion is a significant factor in the lives of sub-Saharan Africans. In addition, it is known that religious groups are important not only for religious activities but also for non-religious ones, such as health and education, in developing countries (Iyer, 2010). Furthermore, the members of religious groups tend to have similar beliefs on health or behaviours (Maselko et al., 2011). Another possible explanation is that religious organisations are more likely to spread prescriptive messages and significantly affect behaviour related to HIV prevention. Thus, lessons shared among religious group members can have a stronger effect on individuals than social norms in neighbourhoods.

2.5 Heterogenous effects by villages

Religious groups and villages do not overlap in the data. Seventy per cent of the religious groups in the main sample comprise members from multiple villages. The effects of religious group members can differ on whether members are in the same village. Therefore, it is meaningful to separately examine the effect of religious group members in the same village and the impact of members in different villages on an individual's choice to obtain their HIV test results. This identifies whether the effects are because of social interactions within religious groups or whether members of the same religious group hold similar views.

Since there exist two dimensions of networks – religious groups and villages – respondents can categorise other respondents into four types according to their religious groups or villages, namely, individuals in the same group and village, those in the same group and other villages, those from other groups and the same village and those from other groups and other villages. I examine the effect of others' decisions to obtain test results in different categories on an individual's choice to obtain their personal test results. Singleton groups of the same religious group and village are dropped. Therefore, the impacts are examined with 670 individuals with at least one peer in the same group and village.

Table 2.7 presents the means and standard deviations for the portion of respondents, the portion of individuals who obtained test results and the portion of those with certain incentive amounts in each category. On average, 1% of all respondents are in the same religious group and the same village, and 1% of all respondents are in the same group and other villages. Since religious groups normally consist of a small number of members, most individuals are in other religious groups and other villages.

Table 2.7: Portion of individuals divided by village and religious group

	N	Mean	SD	N	Mean	SD
Same group	Same village			Other villages		
Portion of all respondents in each category	670	0.01	0.01	670	0.01	0.01
Portion of individuals in each category						
Learned results	670	0.74	0.26	670	0.53	0.39
With incentive 10-100	670	0.38	0.26	670	0.28	0.28
With incentive 110-200	670	0.25	0.24	670	0.18	0.22
With incentive 210-300	670	0.15	0.19	670	0.11	0.17
Other groups	Same village			Other villages		
Portion of all respondents in each category	670	0.02	0.02	670	0.97	0.02
Portion of individuals in each category						
Learned results	670	0.74	0.18	670	0.62	0.15
With incentive 10-100	670	0.39	0.19	670	0.31	0.07
With incentive 110-200	670	0.23	0.14	670	0.21	0.05
With incentive 210-300	670	0.17	0.12	670	0.13	0.03

On average, 53% of religious group members in other villages obtained HIV test results. This outcome is lower than the portion of religious group members in the same village who learned their test results (74%). This result is because of religious groups where all members are in the same village, and the portions of members in other villages who learned results are treated as zero. In addition, monetary incentives are also treated as zero for those with all religious group members in the same village. Thus, the portion of religious group members in other villages assigned monetary incentives larger than zero is relatively smaller than individuals in other categories.

To examine the role of religious and geographical networks in obtaining test results, I expand Equation 2.1 as follows:

$$Y_{ijk} = \alpha + \beta_1 N_{ijk} + \beta_2 N_{ij-k} + \beta_3 N_{i-jk} + \gamma X_{ijk} + e_{ijk} \tag{2.3}$$

Y_{ijk} is an indicator of whether individual i in a religious group j and village k obtained their result. N_{ijk} is the proportion of members in individual i's religious group j and village k who learned their results, N_{ij-k} is the proportion of members in individual i's religious group j and not in village k obtained their results and N_{i-jk} is the proportion of individuals in individual i's village k and not in religious group j that learned their results. X_{ijk} is the vector of controls, and variables are the same as those in Equation 2.1. There are three coefficients of interest. The coefficient on the proportion of religious group members who learned their results in Equation 2.1 was divided to β_1 and β_2 according to whether members are in the same village. In addition, β_3 examines the impact of the proportion of those in other groups but in the same village.

All variables showing the proportion of individuals who learned their test results are instrumented by each type of portion, indicating who is assigned certain amounts of monetary incentives. When estimating Equation 2.3, there are 3 endogenous variables and 3 first stage regressions. The first stage regression for one endogenous variable is as follows:

$$\begin{aligned} N_{ijk} = \alpha &+ \theta_1 IC100_{ijk} + \theta_2 IC200_{ijk} + \theta_3 IC300_{ijk} + \theta_4 IC100_{ij-k} + \theta_5 IC200_{ij-k} + \theta_6 IC300_{ij-k} \\ &+ \theta_7 IC100_{i-jk} + \theta_8 IC200_{i-jk} + \theta_9 IC300_{i-jk} + \delta X_{ijk} + e_{ijk} \end{aligned} \tag{2.4}$$

where instruments are the proportion of religious group members in or not in a village with the values of incentives separated by various cut-off amounts (in Kwacha = US\$ 0.01). Another set of instruments is the proportion of members in other groups and the same village assigned certain amounts of monetary incentives. For example, $IC100_{ijk}$ is the proportion of religious group members in village k who are offered incentives from ten to 100 Kwacha, and $IC100_{ij-k}$ is the portion of group members not in village k who are offered incentives from ten to 100 Kwacha. Likewise, $IC100_{i-jk}$ is the proportion of those who live in village k and not in religious group j who are offered incentives from ten to 100 Kwacha. N_{ij-k} and N_{i-jk} are also instrumented by the same set of instrumental variables.

The regression results are presented in Table 2.8, separated by categories for religious groups and villages. Both the OLS and IV estimates for the impact of religious group members in the same village who obtain test results on an individual to obtain their personal test results are positive and significant. The coefficient is similar to the impact of religious group members combined for all villages in Table 2.6 (0.23). However, the impact of religious group members who are not in the same village and the impact of individuals in other religious groups who obtained test results are not significant. The effects in religious groups that are only significant within the village suggest that peer effects are likely to be driven by social interactions within religious groups rather than similar perspectives of members from the same religious group.

Table 2.9 presents first-stage regression results separated by categories with religious groups and villages. The coefficients of instrument variables are only positive and significant for the dependent variables in the same categories. For example, the portions of religious group members in the same village with incentives have a positive and significant impact solely on the portion of members in the same village and the same group who obtained their test results. This outcome is reasonable because an individual's monetary incentives only affect their decision to obtain test results. The F-statistics of the regressions on the portion of the same religious group members who learned results are large enough to validate the instrument variables. Weak instruments for the portion of individuals who obtained test results in villages that are in other religious groups may not be a serious problem since the corresponding peer effect is not significant.

Table 2.8: Peer effects on learning test results

	Learn results	
	OLS	IV
Portion of those learned results		
Same group & same village	0.86***	0.21*
	(0.03)	(0.13)
Same group & other villages	-0.02	-0.05
	(0.02)	(0.04)
Other groups & same village	-0.03	-0.15
	(0.03)	(0.10)
Monetary incentives		
Non-zero incentives	0.22***	0.30***
	(0.05)	(0.05)
Amounts of incentives	0.00***	0.00***
	(0.00)	(0.00)
N	670	670

SE clustered at the village and religious group level. Covariates: age, age-squared, gender, education, whether has land, marital status, if offered non-zero incentive, the amounts of incentives, whether the distance to the result centre (VCT) was over 1.5km, district, religion, the number of group members, and average distance to locations in VCT zone. *, **, *** indicate significance at the 10, 5 and 1%.

Table 2.9: First stage of peer effects on learning test results

	Portion of individuals learned results		
	Same group		Other groups
	Same village	Other villages	Same village
Same group & same village			
With incentive 10-100	0.40***	0.05	0.08
	(0.08)	(0.07)	(0.06)
With incentive 110-200	0.58***	0.06	0.04
	(0.09)	(0.05)	(0.05)
With incentive 210-300	0.48***	-0.10	0.08
	(0.10)	(0.08)	(0.06)
Same groups & other villages			
With incentive 10-100	-0.05	0.81***	0.03
	(0.06)	(0.06)	(0.04)
With incentive 110-200	-0.09	0.96***	-0.04
	(0.07)	(0.06)	(0.04)
With incentive 210-300	-0.04	0.80***	0.07
	(0.08)	(0.13)	(0.04)
Other groups & same villages			
With incentive 10-100	-0.13	0.01	0.67***
	(0.09)	(0.09)	(0.12)
With incentive 110-200	0.05	-0.13	0.62***
	(0.14)	(0.10)	(0.12)
With incentive 210-300	-0.21	-0.01	0.69***
	(0.14)	(0.11)	(0.15)
N	670	670	670
F	18.06	196.65	7.12

*, **, *** indicate significance at the 10, 5 and 1%. Covariates: age, age-squared, gender, years of education, whether has land, marital status, if offered non-zero incentive, the amount of incentives, whether the distance to the result centre (VCT) was over 1.5km, district, region, the number of group members, and average distance to locations in VCT zone.

2.6 Robustness

Instrument variables with different cut-offs

Instrument variables are the proportions of members with the values of incentives separated by various cut-off amounts, selected by the distribution of respondents and incentive amounts. I conduct the robustness check by changing the cut-offs of incentives for instrument variables. Firstly, I diversify the instrumental variables by dividing incentives to include the same amounts of incentives. Regression (1) includes instrumental variables indicating the portion of members who are assigned incentives divided by every 50 Kwacha. Regression (2) is conducted with instruments generated with monetary incentives divided by 100 Kwacha. Next, I set cut-offs for Regression (3) to distribute the number of respondents as similarly as possible.

Table 2.10: Instruments by different cut-offs for incentive amounts

Regression (1)			Regression (2)			Regression (3)		
Portion	Mean	SD	Portion	Mean	SD	Portion	Mean	SD
10-50	0.18	0.39	10-100	0.37	0.48	10-50	0.18	0.39
60-100	0.18	0.39	110-200	0.25	0.43	60-100	0.18	0.39
110-150	0.08	0.27	210-300	0.15	0.36	110-150	0.08	0.27
160-200	0.17	0.37				160-200	0.17	0.37
210-250	0.07	0.25				210-300	0.15	0.36
260-300	0.09	0.28						

Table 2.11 presents the OLS and IV estimates of the peer effects in religious groups on learning HIV test results with instrumental variables with different cut-offs in Regressions (1), (2) and (3). All regressions with different instrumental variables show a positive and significant impact. Even the sizes of coefficients are similar to the size of peer effects in Table 2.6: 0.23. The first-stage regression results in Table 2.12 present positive and significant impacts of all proportions with monetary incentives for all regressions. F-statistics of all regressions confirm the validity of different types of instrument variables. Therefore, peer effects are robust with instruments divided by different cut-off amounts.

Distance to VCT centre and average incentives

For individuals tested for HIV in 2004, the distances to VCT centres where respondents learn about their HIV test results are also randomly assigned. While monetary incentives can benefit learning test results, the distances to VCT centre constitute costs for learning results. Therefore, the randomised distances could also be the external source of identification to examine the peer effect of religious networks. Group mean distance to a VCT centre is

Table 2.11: Peer effects with different instruments

	Learn HIV test results			
	OLS	IV		
		(1)	(2)	(3)
Portion of members who learned results	0.83***	0.23**	0.21*	0.23**
	(0.05)	(0.11)	(0.11)	(0.11)
N	1045	1045	1045	1045
F	93.89	18.49	17.86	18.40

SE clustered at the village level. Covariates: age, age-squared, gender, years of education, whether has land, marital status, if offered non-zero incentive, the amount of incentive, whether the distance to the result centre (VCT) was over 1.5km, district, religion, the number of group members, and average distance to simulated locations in VCT zone.
*, **, *** indicate significance at the 10, 5 and 1%.

Table 2.12: First stage with different instruments

Incentive	(1)	Incentive	(2)	Incentive	(3)
10-50	0.43***	10-100	0.51***	10-50	0.43***
	(0.07)		(0.06)		(0.07)
60-100	0.58***	110-200	0.62***	60-100	0.58***
	(0.06)		(0.07)		(0.06)
110-150	0.62***	210-300	0.55***	110-150	0.62***
	(0.08)		(0.07)		(0.08)
160-200	0.60***			160-200	0.60***
	(0.09)				(0.09)
210-250	0.56***			210-300	0.54***
	(0.09)				(0.07)
260-300	0.53***				
	(0.09)				
N	1045		1045		1045
F	20.70		32.68		24.72

*, **, *** indicate significance at the 10, 5 and 1%.

around 2km. Furthermore, the value of the incentive itself might be used as an instrument to examine peer effects.

The IV regression results by different combinations of instrument variables in Table 2.14 present positive and significant peer effects. However, the religious group's average distance to the VCT centre and its average value of incentives do not significantly impact the portion of members who obtain test results in the first stage (Table 2.15). Although the joint F

Table 2.13: Instrument variables

	Mean	SD
Portion of group with incentive 10-50	0.18	0.39
Portion of group with incentive 60-100	0.18	0.39
Portion of group with incentive 110-200	0.25	0.43
Portion of group with incentive 210-300	0.15	0.36
Group averages of distances to VCT	2.02	1.01
Group averages of incentive	109.53	47.73

Table 2.14: Peer effects with additional instruments

	Learn HIV test results		
	(1)	(2)	(3)
Portion of members who learned results	0.22**	0.23**	0.23**
	(0.10)	(0.11)	(0.10)
Instrument: Group average distance to VCT	Yes	No	Yes
Instrument: Group average incentives	No	Yes	Yes
N	1045	1045	1045
F	18.34	18.38	18.38

SE clustered at the village level. Covariates: age, age-squared, gender, years of education, whether has land, marital status, if offered non-zero incentive, the amounts of incentives, whether the distance to the result centre (VCT) was over 1.5km, district, religion, the number of group members, and average distance to locations in VCT zone. *, **, *** indicate significance at the 10, 5 and 1%.

Table 2.15: First stage with additional instruments

	(1)	(2)	(3)
With incentive 10-50	0.43***	0.43***	0.43***
	(0.07)	(0.06)	(0.06)
With incentive 60-100	0.58***	0.58***	0.59***
	(0.06)	(0.10)	(0.10)
With incentive 110-200	0.61***	0.61***	0.62***
	(0.07)	(0.16)	(0.17)
With incentive 210-300	0.54***	0.54**	0.56**
	(0.07)	(0.25)	(0.25)
Average distance to VCT	-0.01		-0.01
	(0.01)		(0.01)
Average amount of incentive		-0.00	-0.00
		(0.00)	(0.00)
N	1045	1045	1045
F	24.57	22.82	21.96

*, **, *** indicate significance at the 10, 5 and 1%.

statistics are significant, the coefficients on distances to VCT centres are not significant, suggesting that distances to VCTs do not provide additional variation.

Number instead of the portion of group

It will be worthwhile to examine how the number of group members who learned their HIV test results instead of the portion of the group affects learning test results, allowing for a determination of the direct impact on the number of members who learn results. This specification will examine the effect of the aggregate number of group members as in the local aggregate model, whereas the portion of the religious group who obtain test results in Equation 2.1 is viewed as a social norm in the local average model (Liu et al., 2014). According to Table 2.16, around six religious group members obtained HIV test results on average. The instrumental variables are also substituted with the number of religious group members assigned certain amounts of monetary incentives. Between approximately one and two group members are assigned a certain range of incentives.

Table 2.16: Regressor and instruments with number of group members

	Mean	SD
The number of members who learned test results	6.28	5.37
The number of members with incentive 10-50	1.62	1.82
The number of members with incentive 60-100	1.58	1.51
The number of members with incentive 110-200	2.11	2.04
The number of members with incentive 210-300	1.35	1.58

The results show that both the OLS and IV estimates of peer effects in religious groups are positive and significant in the probability of learning HIV results. The coefficient from the IV regression suggests that, on average, those with one more group member who obtained their test result is 2 percentage points more likely to get their HIV test results. Peer effects of both the portion of the group and the number of members who learned test results imply that policies targeting social norms in religious groups or using social multiplier effects with an individual-based approach can be both valid for enhancing HIV testing.

Table 2.17: Peer effects on number of religious group members

	Learn HIV test results	
	OLS	IV
The number of members who learned test results	0.08***	0.02*
	(0.01)	(0.01)
N	1045	1045
F	48.27	17.66

SE clustered at the village level. Covariates: age, age-squared, gender, years of education, whether has land, marital status, if offered non-zero incentive, the amount of incentive, whether distance to VCT > 1.5km, district, religion, the number of group members, and average distance to locations in VCT zone. *, **, *** indicate significance at the 10, 5 and 1%.

The first-stage regression results present significant and positive effects of the number of members with all types of incentive amounts. F-statistics imply the validity of the instrument variable. Peer effects are robust with the change in variable units to the number of group members who obtained test results.

Table 2.18: First stage with number of group members

	The number of members who learned test results
The number of members with incentive 10-50	0.44***
	(0.07)
The number of members with incentive 60-100	0.64***
	(0.07)
The number of members with incentive 110-200	0.85***
	(0.06)
The number of members with incentive 210-300	0.52***
	(0.08)
N	1045
F	57.56

*, **, *** indicate significance at the 10, 5 and 1%.

2.7 Conclusion

An important goal for HIV prevention is to overcome social stigma, causing more people to get tested and learn about their HIV status. Social networks can diffuse knowledge related to HIV prevention and change the prevailing stigma in sub-Saharan Africa. Religion and religious networks can be useful tools to approach HIV prevention in sub-Saharan Africa since religion significantly affects the lives of people in sub-Saharan Africa, and religious activities can be prescriptive. However, the role of religion is controversial in that many religious leaders taught lessons opposing HIV prevention, such as banning condoms in seeking to discourage immoral sexual activity. In relation to HIV prevention, this chapter examined the effects of religious group members on an individual's behaviour.

The MLSFH data from rural Malawi is used since Malawi represent an environment with poor conditions regarding HIV prevention in sub-Saharan Africa. In 2004, as part of the survey, HIV tests were offered with randomly assigned monetary incentives to encourage individuals to learn about their test results. These randomly assigned incentives were varied to identify peer effects. I used randomly assigned monetary incentives as instrumental variables for the portion of religious groups who obtained HIV test results. The IV estimates generated positive and significant peer effects of religious group members obtaining test results on individuals learning about their personal test results. This effect is stronger than that examined from geographical neighbours. If the regression is repeated by separating religious group members according to whether they are in the same village, only the peer effects of religious group members in the same village are significant and positive. Estimated peer effects on learning HIV test results are robust with instrument variables with different cut-offs for the amounts of incentives, distances to testing centres as additional instruments and the number of religious group members.

This chapter suggests the possibility of using the role of religious networks for HIV prevention in sub-Saharan Africa. The positive peer effects of religious groups on individuals' decisions to learn about their HIV test status in rural Malawi offers a possibility that religion can benefit health policies in developing countries. This result is meaningful in that the behaviour of religious network partners may significantly affect an individual's behaviour in a way that benefits HIV prevention. Policies may consider the possibility of religious networks overcoming social stigma and scaling up health interventions in developing countries. HIV prevention can target religious groups to efficiently increase HIV testing via peer effects with limited resources. Effects larger than those in spatial networks imply that considering religious groups could broaden policy perspectives with better results than interventions through local communities. This result implies that as the measure of the social network becomes closer to the real social network, the network effect becomes larger. The collection of detailed friend and family network data to assess the scale of the effects in these cases can be motivated by this implication. Further studies could seek to understand the channels through which religious networks affect HIV-related behaviour. Also, it is vital to consider the role of religious leaders and religious organisations in the diffusion of HIV-related behaviour through religious networks. Therefore, additional data collection and research methods for networks will allow the development of concrete approaches for HIV prevention using religious networks in sub-Saharan Africa.

Appendix 2A: Sample selection

I estimate whether accepting tests, learning about test results, offering monetary incentives to encourage learning about test results or being HIV positive according to the test affects reporting no religion or the change in religion. Since those responding with no religion or indicating a change in religion between 2004 and 2008 are dropped from the main sample, this analysis will examine the effect on the possibility of being selected for the main sample.

Table 2.19: The impact of HIV tests

	No religion or changed religion		
	(1)	(2)	(3)
Accepted test	-0.00		
	(0.02)		
Learned test results		0.01	0.02
		(0.02)	(0.01)
Amounts of incentives			-0.00
			(0.00)
HIV positive			-0.05**
			(0.02)
N	2186	2009	1986
R^2	0.17	0.16	0.17
F	21.98	17.47	17.32

SE clustered at the village level. Covariates: age, age-squared, gender, years of education, owns land, marital status, district, and religion.
*, **, *** indicate significance at the 10, 5 and 1%.

Table 2.19 presents the OLS estimates for the impact of HIV tests on whether respondents report not aligning with a religion or changing their religion from 2004 to 2008. Accepting tests, learning test results and amounts of incentives do not significantly affect reporting no religion or change in religion between 2004 and 2008. This outcome implies that accepting getting tested or learning about test results is not critically related to the sample selection process.

Chapter 3

Peer effects on subjective expectations of HIV

3.1 Introduction

Although numerous medical methods for HIV prevention are available, HIV is still one of the most serious health problems in sub-Saharan Africa. Achieving a substantial decrease in HIV prevalence is difficult because HIV prevention requires understanding risks with relevant information and changing behaviour based on subjective expectations regarding HIV infection (Gong, 2014; Schaefer et al., 2020). Subjective expectations significantly influence health outcomes by inducing the adoption of preventive health behaviours and avoidance of risky ones (Delavande and Kohler, 2009). In developing countries, where people lack health services and education, social networks are potentially critical sources of information affecting health beliefs. Therefore, it is important to understand peer effects on subjective expectations related to HIV as a mechanism to effectively conduct health interventions for HIV prevention.

Studying peer effects on subjective expectations of HIV in sub-Saharan Africa reveals important aspects of HIV prevention, including information provision and social learning.

Knowing others' subjective beliefs about their HIV infections can provide information on HIV prevalence for those who lack knowledge of the severity of HIV in their communities. Furthermore, social learning can support individuals in understanding the salience of HIV, which is likely to be unknown because HIV prevalence and symptoms are not visible (Linnemayr, 2018). Such information could lead many individuals to prioritise HIV prevention and change their behaviour.

Estimating the peer effects is related with social learning, whereby social networks in rural villages could be perceived as playing a significant role in forming health beliefs in developing countries. Positive peer effects could show that social learning exists and individuals in rural villages are likely to conform to the subjective beliefs of the majority in their villages. Villages can be a meaningful unit because residents share similar environments and social communities. Thus, subjective beliefs on HIV prevalence and risks in villages might affect the formation of subjective beliefs among village neighbours through social learning. In addition, peer effects in villages may indicate that beliefs could be affected not only by those who are directly connected but also by those linked indirectly in villages.[1] However, not many studies explicitly examined the effects of network partners on the perceived risk of HIV infection because of the lack of relevant data about subjective expectations regarding HIV in developing countries and the difficulties of identifying the effects within social networks. In this chapter, I estimate peer effects on the subjective expectation of contracting HIV in villages using data from rural Malawi. Studying peer effects on subjective expectations about HIV infection is a new approach compared to the literature of HIV prevention strategies where subjective expectations are studied in terms of factors affecting health behaviour. This study explores the possible role of social networks on HIV prevention using subjective expectations. The approach also adds on the literature related with social learning in terms of health belief about one's own HIV infection.

[1] Peer effects on subjective expectations in religious groups, as an extension from the previous chapter, may not reflect the full effect, including those through indirect connections. Indeed, peer effects on subjective expectations were not significant in religious groups.

I overcome the difficulties of estimating peer effects on subjective expectations by using rich data from surveys in rural Malawi, which collected data on the subjective expectations of HIV infection. The data also include randomised experiments providing the source of variation for identification. Subjective expectations are collected by asking respondents to allocate up to ten beans to indicate the likelihood of their own HIV infection. This method allows the advantage of being observable and relatively attractive to respondents (Delavande and Kohler, 2012). However, researchers face challenges when examining peer effects because it is difficult to distinguish the peer effects of subjective expectations from other effects. The correlation in expectations among peers might be caused by unobserved characteristics due to the endogenous formation of networks, common shocks in networks and similar characteristics of peers. Therefore, to disentangle the peer effects based upon the subjective expectations of neighbours in villages, I use the variation of randomised monetary incentives offered from the HIV test before collecting subjective expectations. Using randomly assigned monetary incentives as exogenous instruments provide the source of variation to estimate the causal effects of others' subjective expectations on an individual's subjective belief about HIV infection.

I examine peer effects in villages on an individual's subjective expectation of HIV infection by regressing subjective expectations on the average village expectations, instrumented by randomised monetary incentives to learn about test results. The estimates of instrumental variable regressions find significant and positive peer effects on one's subjective likelihood of both current HIV infection and infection in the future. This finding suggests that subjective expectations related to HIV can be formed by network effects within villages. Social learning can be performed by conforming to the majority of the village and forming one's subjective expectations of HIV infection. Through these peer effects, information provision on HIV prevalence and salience in villages can be a key to changing health behaviour for HIV prevention in developing countries. In this analysis, I identify peer effects in subjective expectations rather than behavioural changes caused by learning test results. The identification assumption is that one's randomly assigned incentives only affect one's subjective expectations directly by learning one's own test results. This assumption for identification is reasonable since the incentives only affected the likelihood of learning one's own test result, and the incentives were assigned privately. Therefore, the effects would operate through updating one's own beliefs. This assumption implies that one's behaviour is not directly affected by others' incentives to learn test results. The effects could change behaviour eventually, but the first step is the change in subjective expectations. Thus, other effects aside from subjective expectations triggered by test results are not identified in this analysis.

Since learning about HIV test results may correct subjective expectations of HIV infection, examining peer effects by whether peers learn about test results is meaningful to understanding the role of HIV testing on peer effects. Thus, I use the marginal treatment effect to examine peer effects, depending on whether village neighbours learned test results and have a low subjective likelihood of HIV infection. Marginal treatment effect (MTE) parameters show that peer effects of those who learned HIV test results and have subjective likelihood lower than median subjective expectations are significant and larger than others. This finding not only reveals heterogenous peer effects depending on the subjective likelihood of neighbours but also emphasises that peer effects can be stronger with HIV testing. The need to promote HIV testing is implied as a relevant policy to take advantage of peer effects on subjective expectations.

3.2 Literature

3.2.1 Subjective expectations about HIV

Many works of literature have studied the relationship between perceived HIV risk and sexual behaviours. Gerrard et al. (1996) review the literature and point out that many studies conclude that those who engage in risky sexual behaviours are more likely to have a high subjective likelihood of being infected with HIV. Therefore, perceived HIV risk is considered crucial in intervening in health behaviour. For instance, Schaefer et al. (2020) estimated associations between changes in perceived HIV risk and changes in condom use. Using data on sexually active adults in Zimbabwe, the researchers found an increase in perceived HIV risk associated with a higher likelihood of more condom use.

Paula et al. (2014) analysed data from surveys of the Malawi Diffusion and Ideational Change Project (MDICP) and investigated the relationship between subjective beliefs about individuals' personal HIV status and risky sexual behaviours. The policy implication of this result is that providing those infected with HIV with accurate information on their HIV status may lead them to reduce their likelihood of engaging in risky sexual behaviour. Anglewicz and Kohler (2009) pointed out that individuals overestimate their likelihood of being HIV positive when they evaluate their own and others' sexual histories. Subjective expectations about HIV infection of rural Malawians are compared with the actual HIV biomarker results and a lot of respondents showed a tendency to overestimate their infection risk. The authors pointed out that the overestimation of HIV status is likely to be caused by the overestimation of HIV transmission probabilities.

Acquiring information on HIV risk is also claimed to change behaviour. In this regard, Dupas (2011) used the data from a randomised experiment in Kenya and studied the role of providing information on risk reduction strategies for HIV prevention. The result showed that the HIV curriculum, including risk information, reduced unprotected sex among teenage girls. Furthermore, increased reporting of sexual behaviour among teenage boys implied that the responses provided by teenage girls were indicative of them substituting away from riskier partners who are old. Therefore, HIV interventions can target subjective HIV risks to cause behavioural changes by providing relevant information on HIV and risky behaviours.

Studies have claimed the importance of data on subjective expectations. Accordingly, Attanasio (2009) insisted on the need to collect data on expectations related to important decisions because of the importance of understanding how subjective expectations affect behaviour in the developing world. Therefore, Attanasio (2009) said that data on beliefs should be collected systematically in surveys. Moreover, Delavande and Kohler (2009) maintained that subjective expectations are critical to inferring the determinants of behaviours. Therefore, the authors implied that data on health behaviour without information on subjective beliefs do not allow researchers to determine policy implications. For example, a decision to not use condoms might be made not only for individuals with a high subjective risk of HIV infection who do not like regularly using condoms but also for those who do not have a preference regarding condom use, instead preferring to believe in having the same serostatus with their partners. Furthermore, a person who does not trust the function of condoms may not choose to use them. Thus, data on condom use itself may not provide the optimal direction for policy implications, such as the provision of information on condom use, educating on safe sexual behaviour or scaling up testing. The gains from longer life expectancy might be limited due to the high mortality levels in sub-Saharan Africa. Thus, adopting safe sexual behaviour might not be optimal for those in sub-Saharan Africa when comparing costs (financial expenses and the costs of giving up pleasures) and benefits (longer life expectancy and healthier lives) (Oster, 2007; Delavande and Kohler, 2016).

3.2.2 Peer effects in subjective expectations

Another important line of literature is why and how social networks can influence beliefs. Algan et al. (2019) studied the effect of friendship interactions on political opinions with the network of first-year students at Science Po. The difference in political opinion is reduced by a friendship link. The study successfully estimated the causal effect of networks on personal opinions by using the variation of exogenously connected pairs of students. Many studies theoretically and empirically investigated belief formation with social learning models in networks. The two most widely used models used to explain how social networks influence individuals' beliefs are Bayesian learning and naive (DeGroot) learning. Bayesian learners observe their network partners' beliefs and form their own repeatedly and simultaneously, whereas naive learners follow the majority of their network partners' opinions. Mixed opinions exist on which model explains the formation of beliefs (Chanderasekhar et al., 2020; Grimm and Mengel, 2020).

Researchers indicate that social interactions have played an essential role in forming individuals' perceptions of their risk of HIV infection. Social interactions are also mentioned as a critical factor for evaluating the benefits and costs of behavioural changes such as the adoption of condom use and increased marital fidelity. Kim (2016) used the school-based survey data of Malawian adolescents to estimate peer effects among classmates on an individual's probabilistic expectation of being infected with HIV. The positive and significant peer effects on an adolescent's subjective likelihood of HIV infection are found. This result might be relevant to policies using social interactions to affect the subjective expectations of adolescents for HIV prevention.

Social networks are important for affecting knowledge and attitudes regarding HIV by diffusing knowledge and influence. A qualitative study conducted by Buhler and Kohler (2003) in the South Nyanza District in Kenya shows that the perceived risks of HIV infection depend on the common perceived risks within personal communication networks, affected both by the content of AIDS-related communication and the network structure in communication. Smith and Watkins (2005) suggested that worry about HIV is associated with the extent to which social network partners worry about HIV. Investigating the survey data of rural Malawians, the researchers determined the effect of respondents' conversational networks on an individual's anxiety surrounding HIV. Thus, perceived risks of HIV infection can indirectly indicate behaviour change.

Several studies mentioned the difficulties of analysing peer effects on subjective expectations. Attanasio (2009) argued that collecting data on expectations has not been an easy task in developing countries. This challenge has arisen because many respondents in developing countries have less opportunity to receive formal education and are unfamiliar with probabilities. Inaccurate or inconsistent responses are likely attained from surveys asking what level respondents anticipate for the probabilities of events in developing countries.

Kim (2016) addresses the low reliability of responses about subjective expectations related to probabilities by suggesting that local contexts are used and questions asked differently. For example, a ruler graded from zero to 100 was used to help respondents understand their probabilistic expectations of certain events. Delavande and Kohler (2012) collected subjective expectations by requiring respondents to allocate up to ten beans to indicate the likelihood that an event, such as HIV infection, would happen. This method allows the advantage of being observable and relatively attractive for respondents to express their subjective expectations. These data are found to produce subjective expectations consistent with the basic characteristics of probability theory. Thus, a well-designed approach enables data collection on subjective beliefs relevant for analysis in developing countries.

The endogeneity of the network has been mentioned as a key challenge when identifying peer effects. Regarding subjective expectations, the effects of peers' beliefs are difficult to disentangle from the effects of their characteristics and correlated effects. The study analysing the effect of conversational networks on one's worry about HIV mentioned the bias from network endogeneity as a limitation for the analysis (Smith and Watkins, 2005). Kim (2016) used the proportion of classmates with family members who died of HIV as an instrument to address this identification problem. The instrument is considered valid since many students tend to report the HIV-related deaths of distant family members. Therefore, classmates would not necessarily know about these deaths and will only be affected by their classmates' HIV infection expectations. However, HIV infections among distant family members can also directly influence the subjective expectations of classmates, depending on the relations or distance among classmates' families. I address the challenge by using randomised monetary incentives to encourage learning about test results as the source of exogenous variation to identify peer effects on subjective expectations.

3.3 Background and data

3.3.1 Background and survey design

To estimate the peer effects on subjective expectation, I use the data from the MLSFH, which was collected in three rural districts in Malawi from 1998 to 2012. This study provides longstanding data about HIV, sexual behaviours, and families, studied over seven survey rounds. The sample for MLSFH was selected to represent the rural Malawi population. The MLSFH data can be helpful since the dataset collected rich information on the subjective expectations of individuals in rural Malawi. Thus, many researchers conducted studies related to subjective expectations using the MLSFH dataset, producing various results. For instance, Delavande and Kohler (2009) examined the impact of individuals' expectations about future HIV status on risky sexual behaviour with probabilistic expectation data from MLSFH data. Delavande and Kohler (2012) used subjective belief data from the MLSFH to study the effect of learning HIV status on an individual's subjective expectations of being infected with HIV and subsequent risky behaviour.

I use data from the 2004 and 2006 survey rounds that include information on HIV testing with randomly assigned incentives to learn test results and subjective expectations on HIV infection. In 2004, respondents were offered free tests for HIV status. Respondents who accepted to get tested were offered randomly assigned monetary incentives to learn about their test results. Next, a survey was conducted in 2006, and subjective expectations on various events, including HIV infection, were collected. Although both surveys in 2004 and 2006 collected data on subjective beliefs, the survey in 2006 suggested a way to collect subjective expectations with a numerical measure. This initiative provides more precise measures of subjective beliefs compared to the coarse categories in the 2004 survey. The questions in the 2004 survey are 'In your opinion, what is the likelihood that you are infected with HIV now?' and 'In your opinion, what is the likelihood that you will become infected with HIV in the future?' Possible answer choices were no likelihood, low likelihood, medium likelihood and high likelihood. These coarse categories are simple to provide explanations to respondents and easy to collect and process. However, low, medium and high likelihood might be very different among individuals and regions. Furthermore, collected information is limited due to less variations in categories. Because the scales suggested from the 2004 survey are verbal, the degrees of likelihood may not be comparable across individuals (Delavande and Kohler, 2009).

The 2006 survey introduced a new method to collect expectation data from respondents. In a separate section of the questionnaire, respondents were asked several questions about the chance that certain events were going to happen. Respondents answered by putting some beans out of ten beans in a cup to express their opinion of the likelihood of certain events going to happen. In this regard, one bean meant one chance out of ten that an event was likely to happen. Putting no bean meant that the respondent thought that the event would certainly not happen. In addition, the survey offered explanations with examples of what each number of beans represents. For example, five beans meant that the likelihood of an event happening is the same as the likelihood that it does not occur. Six beans represented that the event is slightly more likely to occur than not. This segmentation of responses and detailed explanations provided flexibility for respondents to answer. Therefore, exhaustive information on subjective expectations is offered from data in the 2006 survey. Eliciting probabilistic expectations used a numeric scale instead of a verbal scale, and responses could be consistently converted to probabilities across different answers (Delavande and Kohler, 2009). The process was interactive in that respondents were provided with detailed explanations and were open to clarifying questions. Simple questions were preceded to evaluate how respondents understood the concept of probability, and adjustments were made during the survey.

Delavande and Kohler (2009) studied the accuracy of reported numerical beliefs about HIV infection and discovered that the probability assessments on being HIV positive were substantially well-calculated compared to HIV prevalence rates in the local community.

Furthermore, the method of collecting the data was efficient in that the collection method was visual, relatively intuitive and fairly engaging for respondents. Delavande and Kohler (2009) argued that the non-response rate was significantly lower than others in surveys asking about subjective expectations in developed countries. In addition, most respondents responded consistently with the properties of probability theory. The variation of responses according to demographic characteristics was not substantially different from the expected directions. These characteristics suggested that the subjective expectation data in the 2006 MLSFH survey reflected the ability of individuals in developing regions to provide valid responses about their beliefs about HIV infection.

3.3.2 Sample selection

The main sample is individuals who accepted an offer to get HIV tested in 2004 and interviewed in 2006 with no missing data in monetary incentives for HIV testing, test results, village information and subjective expectation. Among 1,617 respondents who were offered HIV tests in rural Malawi, 92% accepted to get tested for HIV. From 1,479 individuals who accepted the HIV test, 1,448 respondents were chosen as the main sample without missing data. Since peer effects are examined within villages, not only those without village information but also respondents who are the only resident in their village in the sample (living in a singleton village) are dropped from the main sample. To check possible selection problems, I tested whether the selected and dropped samples were similar in terms of observed characteristics by using pairwise t-tests for selection procedures. Table 3.1 presents the results of the pairwise t-tests to compare characteristics (gender, age, marital status, education, land ownership and HIV status). In this regard, there are no significant differences in all characteristics between dropped and selected samples.

Table 3.1: Pairwise t-tests for sample selection

	Accepted test			No missing variable		
	Drop	Select	P-value	Drop	Select	P-value
Male	0.40	0.38	0.70	0.36	0.38	0.79
Age	36.43	36.66	0.88	35.04	36.70	0.45
Married	0.90	0.88	0.54	0.86	0.88	0.68
Education	3.12	3.29	0.50	3.46	3.29	0.75
Owns land	0.79	0.79	0.91	0.79	0.79	0.92
HIV+				0.05	0.04	0.79
N	138	1479		31	1448	

3.3.3 Data

Table 3.2 presents descriptive statistics of the main sample consists of 1,448 individuals. 38% of the main sample are male. Respondents selected for the main sample are, on average, 37 years old with an average of three years of education. Additionally, 88% of the main sample are married, and 79% own land. According to health beliefs collected in 2004, which may be correlated with subjective expectations of HIV infection, most respondents consider themselves healthy. Almost half of the respondents considered their health conditions to be better than two years ago. Only 6% believed their health worsened compared to two years ago. The respondents in the main sample live in 112 villages total. Villages consist of 13 members on average, and the number of residents in each village varies from two to 85 members.

Moreover, 74% of the main sample learned their test results. Individuals who accepted HIV tests were offered randomly assigned monetary incentives. The value of monetary incentives varies from zero to 300 Kwacha. The maximum value of incentives, 300 Kwacha, is approximately US$ 3 (Pauw et al., 2013), which is similar to the average two days' wages for rural Malawians (Fedor et al., 2015). The value of the incentive offered to those in the main sample is 105 Kwacha (approximately US$ dollar) on average. The average monetary incentive to learn about HIV test results is less than one day's wage for rural Malawians. Thornton (2008) argued with this data that fear of or stigma surrounding HIV testing could

Table 3.2: Descriptive statistics

	N	Mean	SD
Age	1448	36.70	12.63
Male	1448	0.38	0.49
Years of education	1448	3.29	2.88
Owns land	1448	0.79	0.40
Married	1448	0.88	0.32
Health belief			
My health is good	1448	0.94	0.24
My health is better than that of two years ago	1448	0.52	0.50
My health is worse than that of two years ago	1448	0.06	0.23
HIV test in 2004			
Learned test result	1448	0.74	0.44
Incentive to learn result	1448	105.28	94.15
Non-zero incentive	1448	0.77	0.42
HIV+	1448	0.04	0.19
Any subjective likelihood of HIV+	1270	0.36	0.48

[1] The main sample consists of individuals who accepted an offer to get tested without missing data. Although subjective expectation in numerical measures are available for all individuals in the main sample, responses in coarse category is available for 1270 individuals only.

be easily overcome by small cash incentives. In this regard, 77% of the main sample received a non-zero incentive. As a result of HIV tests, 4% of the main sample turned out to be HIV positive, which is substantially lower than the portion of respondents who believed there is any likelihood of HIV infection (36%). This result implies that respondents overestimated the likelihood of HIV infection.

3.3.4 Subjective expectations data

Responses in the 2006 survey regarding the subjective likelihood of being infected with HIV is around one in a scale of zero to ten. This outcome is 0.11 when divided by ten, which is very low. It is lower than the average response asked by coarse degrees of subjective likelihood (28%). In addition, this is lower than the coarse subjective likelihood of HIV infection provided in responses before the HIV test in 2004 (36%). The subjective belief of being infected with HIV decreased after the test in 2004. The subjective likelihood of future HIV infection on a zero to ten scale is around one when responses were provided by allocating beans among ten beans. This outcome is similar to current HIV infection on a zero to ten scale. When responses are collected by a verbal scale, the portion of respondents who responded that there is any likelihood of HIV infection in the future is 0.52, which is much higher than the subjective likelihood of future HIV infection collected on a numerical scale. This outcome is also higher than the subjective belief about current HIV infection collected by a verbal scale (0.28).

Table 3.3: Subjective likelihood of HIV infection

		N	Mean	SD
Current infection	In 0-10 scale / 10 [1]	1448	0.11	0.20
	Any likelihood [2]	1439	0.28	0.45
Future infection	In 0-10 scale / 10	1346	0.11	0.21
	Any likelihood	1388	0.52	0.50

[1] Subjective likelihood that the respondent is currently infected with HIV responded in 0-10 scale divided by 10.

[2] =1 if responded there is any likelihood that the respondent is currently infected with HIV as opposed to no likelihood.

The subjective likelihood of HIV infection might differ between those infected with HIV and those who are not. Subjective expectation indicators investigated in 2006 are presented separately by actual HIV status from the 2004 HIV test. For those who tested positive, the likelihood of current infection is far from one. This interesting result might be due to those who do not believe the test result or do not respond, reflecting their actual beliefs. A higher portion of those who tested HIV positive responded higher subjective likelihood of being infected with HIV for both current and future infection than those who tested HIV negative. If the subjective likelihood is collected in numerical measure, the responses of those infected with HIV are more than doubled from the responses of those who tested negative. Even those who tested negative have a higher subjective likelihood of future HIV infection than the likelihood of current infection. However, subjective likelihood collected on a zero to ten scale is similar between current and future infection for those who tested negative.

Table 3.4: Subjective expectations by HIV status

	HIV negative			HIV positive		
	N	Mean	SD	N	Mean	SD
Any likelihood of current infection	1382	0.27	0.45	57	0.56	0.50
(In 0-10 scale / 10)	1391	0.10	0.19	57	0.27	0.32
Any likelihood of future infection	1331	0.51	0.50	57	0.79	0.41
(In 0-10 scale / 10)	1301	0.11	0.20	45	0.31	0.35

Although there are not many respondents who tested HIV positive, it is surprising to observe some individuals among them who responded that there is no likelihood of HIV infection. Forty-four per cent of those who tested HIV positive still responded that there is no likelihood of being infected with HIV. Furthermore, 21% of those who tested HIV positive still believe there is no likelihood of being infected with HIV even in the future. Possibly, individuals may not trust the HIV test result or forget about their test results. Another plausible explanation is that respondents know they are HIV positive but are embarrassed to respond truthfully in front of the interviewer (Delavande and Kohler, 2016). This situation is likely because HIV is related to psychological barriers, such as fear or stigma, which is particularly applicable to sub-Saharan Africa.

3.4 Estimation

3.4.1 Identification strategy

In measuring the peer effects of neighbours in the village on subjective expectations of one's own HIV status, a necessary procedure is to find a strategy to address the difficulties of identifying peer effects. It is complicated to distinguish causal effects from peers since an individual's behaviour can be derived from various types of effects. These are the influence of exogenous peer effects, the behaviour of individuals in the network, and the fact that individuals in the same network behave similarly due to their similar characteristics (Manski, 1993). In this estimation, the respondent's subjective likelihood of HIV infection can be driven by not only the subjective expectations of others in the same village but also the characteristics of neighbours in the village. Furthermore, peer effects are difficult to disentangle from correlated effects since residents in the same village live together by choices common in the village, or common shocks might have occurred at the village level.

Figure 3.1: The order of events

In the experimental design of the MLSFH survey in 2004, monetary incentives were randomly assigned to encourage individuals to obtain their 2004 HIV test results. Learning about test results, inspired by monetary incentives, affected subjective expectations of HIV infection collected in the 2006 survey. Since randomly assigned monetary incentives are exogenous, it is plausible to use monetary incentives as the source of variation in the estimation. In this regard, the order of events is as follows: In the 2004 survey round, the data with the coarse measure of the likelihood of being infected was collected. After the survey, an HIV test was offered, and individuals who accepted the offer were tested. Those tested were offered randomised monetary incentives to encourage them to learn about their test results. Later in 2006, another survey was initiated, which included collecting data on subjective expectations. Peer effects are estimated for the effect of these subjective expectations on others' subjective expectations.

Table 3.5: The effect of learning test result on subjective likelihood of HIV infection

	Likelihood of HIV infection	
	Infected now	Infect in the future
Learn test result	-0.03***	-0.02*
	(0.01)	(0.01)
N	1448	1346
R^2	0.12	0.13

*, **, *** indicate significance at the 10, 5 and 1%. Covariates: age, age-squared, gender, education, owns land, married, distance to testing centre > 1.5km, district, religion, distance to locations in VCT zone, subjective health and HIV status

Table 3.5 presents OLS estimates for the effect of learning about HIV test results on the subjective likelihood of an individual's HIV infection on a zero to ten scale divided by ten for both current and future infection. Learning about test results significantly and negatively affects the subjective likelihood of HIV infection for both current and future infection. Negative effects might be driven by those who tested HIV negative, amending their subjective likelihood of HIV infection downwards. In addition, monetary incentives motivated individuals to learn about their test results, which eventually reduced the subjective likelihood of the individual becoming infected with HIV. Therefore, I use these incentives as instrument variables to examine the peer effects of subjective expectation.

3.4.2 Model

I examine peer effects in the village on an individual's subjective expectation of HIV infection by regressing subjective expectations on average village expectations, instrumented by randomised monetary incentives to encourage learning about test results as follows:

$$Y_{ij} = \alpha + \bar{Y}_{ij}\beta + X_{ij}\gamma + e_{ij} \tag{3.1}$$

where Y_{ij} is the subjective likelihood of HIV infection of individual i in village j collected on a zero to ten scale divided by ten. Thus, this variable takes values between zero and one. This result can reflect the subjective likelihood of current HIV infection or infection in the future. $\bar{Y}_{ij}$ is the leave-out mean of the likelihood of subjective HIV infection. X_{ij} is the vector of controls, which will be explained in the following paragraphs. The main coefficient, β, shows the peer effects of the average subjective likelihood of HIV infection in an individual's village. Using the village mean as an explanatory variable to estimate peer effect reflects how an individual tends to conform to or imitate their neighbours. The numerical indicator on a zero to ten scale reflects the degrees of expectation and provides beliefs consistent with basic properties of probability theory (Delavande and Kohler, 2012).

The vector of controls includes age, age-squared, gender, years of education, a dummy on whether a respondent owns land, marital status, district, religion, the distance to the VCT centre, where respondents learn about their test results, a dummy =1 if the distance to VCT is over 1.5km, district, the average distance to simulated locations in the VCT zone[2], subjective health status, HIV status from the test and the village leave-out means of these covariates. These variables are included to control the effect of individuals' demographics and locations. Village leave-out means of covariates are included to net out the effect of other village neighbours' characteristics from peer effects. Subjective health status includes dummy variables indicating whether respondents reported their health as being good, better than two years ago or worse than two years ago. These variables can control self-reported health that should correlate with HIV expectations (Delavande and Kohler, 2009).

[2]The average distances to simulated locations in VCT, as calculated by Thornton (2008), is included as a control for the distance to VCT and the location of respondents in each VCT zone since the location of VCT is chosen randomly in each zone. This control variable was included in several studies using the same MLSFH data related to the 2004 HIV tests (Thornton, 2008b; Godlonton and Thornton, 2012).

3.4.3 First stage

The first stage regression is as follows:

$$\bar{Y}_{ij} = \alpha + IC50_{ij}\theta_1 + IC100_{ij}\theta_2 + IC150_{ij}\theta_3 + IC200_{ij}\theta_4 + IC250_{ij}\theta_5 + IC300_{ij}\theta_6 + X_{ij}\delta + e_{ij} \quad (3.2)$$

where $IC50_{ij}$ indicates the portion of the village with incentives from ten to 50, $IC100_{ij}$ indicates the portion of the village with incentives from 60 to 100, $IC150_{ij}$ means the portion of the village with incentives from 110 to 150, and so on.

Instruments are the set of splines composed of the proportion of individuals in each village with the amounts of incentives separated by various cut-off amounts (in Kwacha = 0.01 USD). Ten Kwacha is the minimum value of the incentive, and 300 Kwacha is the maximum. Each variable is generated with cut-off amounts of every 50 Kwachas (approximately 50 cents in USD). I use this set of splines as the set of instruments to allow flexibility in the relationship between instruments and variables, such as non-linear relationships. The findings are robust to other formulations. Other studies analysing the same dataset with randomised monetary incentives, including Godlonton and Thornton (2012) who estimated peer effect on learning HIV test results, also used the set of splines consisting of certain monetary incentives as instrumental variables.

Table 3.6 shows the OLS estimates for the effects of monetary incentives on an individual's subjective likelihood of being infected with HIV now and in the future. These findings are the reduced form effect of the monetary incentives on individuals' probability of being infected with HIV. Interestingly, those assigned monetary incentives from 160 to 200 Kwacha have a higher subjective likelihood of HIV infection. On the other hand, those assigned monetary incentives between 210 and 250 Kwacha report a lower subjective likelihood on the zero to ten scale. This result is intriguing because it reflects the possibility of heterogeneous effects depending on the value of monetary incentives. This reduced form of

Table 3.6: Reduced form: Effect of incentives on subjective likelihood

	Subjective likelihood	
	Infected now	Infect in the future
Incentive 10-50	0.00	0.00
	(0.02)	(0.02)
Incentive 60-100	0.01	0.02
	(0.02)	(0.02)
Incentive 110-150	0.01	0.01
	(0.02)	(0.02)
Incentive 160-200	0.04**	0.04**
	(0.02)	(0.02)
Incentive 210-250	-0.04*	-0.05*
	(0.02)	(0.03)
Incentive 260-300	0.01	0.00
	(0.02)	(0.02)
N	1448	1346
R^2	0.12	0.14

*, **, *** indicate significance at the 10, 5 and 1%. Covariates: age, age-squared, gender, education, owns land, married, distance to testing centre > 1.5km, district, religion, distance to locations in VCT zone, subjective health and HIV status

regression supports the need to allow flexibility for the relations between instrumental and explanatory variables in peer effects estimation. Larger incentives tend to carry a greater inducement for individuals to learn about their test results. Therefore, it is reasonable to expect that those assigned large incentives, higher than 200 in this case, are more likely to learn about their test results and eventually have a lower subjective likelihood. Using the spline of functions with different types of incentive values could reflect non-linear and heterogenous effects by incentive values shown in this result.

Table 3.7 presents the result of the first-stage regression. The first-stage estimates show that the portions of the village with a certain quantum of incentives affect the village-average subjective likelihood of HIV infection for both current and future infection. This outcome implies instrument relevance since instrument variables significantly impact independent variables. The F-statistics is around 12 for the average likelihood of current infection and around 19 for the average likelihood of infection in the future, both larger than the suggested cut-off (ten) for weak instruments (Stock and Yogo, 2005). In addition, different signs and significance of coefficients for different instruments validate the form of instrument variables.

Table 3.7: First stage for peer effects on subjective likelihood

	Mean (Subjective likelihood)	
	Infected now	Future infection
Incentive 10-50	0.04**	0.06***
	(0.02)	(0.02)
Incentive 60-100	0.02	0.05***
	(0.02)	(0.02)
Incentive 110-150	-0.03	-0.10***
	(0.03)	(0.03)
Incentive 160-200	0.11***	0.10***
	(0.02)	(0.02)
Incentive 210-250	-0.09***	-0.11***
	(0.03)	(0.03)
Incentive 260-300	-0.02	-0.07***
	(0.02)	(0.03)
N	1448	1346
F	12.07	18.91

*, **, *** indicate significance at the 10, 5 and 1%.
Covariates: age, age-squared, gender, education, owns land, married, district, religion, distance to testing centre, distance $>$ 1.5km, distance to locations in VCT zone, subjective health, HIV status, and village means of covariates.

The first-stage regression results are interesting in that the direction of effects is not the same for the village portions with different amounts. For example, for the mean subjective likelihood of current HIV infection, the segment of the village with incentives from ten to 50 and from 160 to 200 increases its mean subjective likelihood, whereas the segment of the village assigned incentives from 210 to 250 negatively affected the mean subjective likelihood. This outcome is consistent with the negative coefficient from the regression estimating the effect of learning test results on the subjective likelihood of current infection. Therefore, higher average amounts of monetary incentives cause more individuals in villages to learn about their test results and eventually, on average, reduce the subjective likelihood of HIV infection in villages. These features suggest if monetary incentives are assigned, individuals might be motivated to learn about their test results and change their beliefs on their HIV status even though there exists a stigma around them. Furthermore, different values of incentives have heterogenous effects, and higher incentives can be more effective in overcoming social constraints to obtaining test results and changing subjective expectations.

3.4.4 Results

Table 3.8 shows the OLS and IV estimates for the peer effect within villages. For both OLS and IV regressions, peer effects are positive and significant for both the subjective likelihood of current and future infection. These estimates imply that a one-percentage-point increase in the average subjective likelihood of an individual's village significantly increases their subjective likelihood of current HIV infection by 0.59 percentage points. Furthermore, a one-percentage-point increase in the village's average subjective likelihood increases an individual's subjective belief of future HIV infection by 0.65 percentage points. An individual's subjective expectation of HIV infection is significantly affected by the subjective expectations of neighbours in their village.

Table 3.8: Peer effects on subjective expectations of HIV infection

	Subjective likelihood			
	Infected now		Future infection	
	OLS	IV	OLS	IV
	(1)	(2)	(3)	(4)
Village mean	0.19**	0.59*	0.22***	0.65**
	(0.08)	(0.35)	(0.08)	(0.27)
N	1448	1448	1346	1346

*, **, *** indicate significance at the 10, 5 and 1%.
Covariates: age, age-squared, gender, education, owns land, married, district, religion, distance to testing centre, distance > 1.5km, distance to locations in VCT zone, subjective health, HIV status, and village means of covariates.

An interesting feature of the result is that estimates from IV regressions are larger than OLS estimates for both regressions on the likelihood of current and future infection. IV estimates larger than OLS estimates are different from the expected direction of bias. Due to the endogeneity of OLS estimates for network effects, IV estimates are expected to be smaller than OLS estimates. Several reasons can explain this, according to studies by Card (2001), Dickson (2009) and Oreopoulos (2006). One possible explanation is downward bias in OLS estimates due to measurement error. For example, measurement errors in subjective expectations may cause attenuation bias (Griliches, 1977). However, Dickson (2009) argued that survey measurement errors could only explain a small amount of downward bias. Another reason might be that IV estimates represent the local average treatment effect (LATE). Peer effects might be identified for those with neighbours in villages whose subjective beliefs are affected by learning test results because of monetary incentives. Thus, the peer effects of compliers might be larger than the average effects of the entire population. Card (2001) mentioned IV estimates could be larger than those of OLS might reveal higher-than-average effects on outcome for a subset of the population. LATE may provide meaningful information when considering policy purposes since it examines the effect on the individuals in question. Therefore, LATE could be a more relevant measure than average effects on the whole population (Dickson, 2009).

3.5 Marginal Treatment Effect

Peer effects from the IV estimation represent the impact on those with village neighbours who obtained their test results due to monetary incentives. IV estimates larger than OLS estimates infer that LATE from the IV estimation is indicative of compliers face larger peer effects than others. Since subjective likelihood is affected by whether individuals learned that their test results, it is worth analysing the MTE and examining peer effects depending on peers' subjective likelihood compared to others. This analysis will help understand the mechanism from HIV testing to peer effects. Larger effects of those who learned test results could be meaningful for policies to affect subjective beliefs through peer effects by scaling up HIV testing.

The MTE is suggested as an approach to recover economically interesting parameters in models with heterogeneous treatment effects. MTE is the mean response of individuals to treatment at a margin. It allows using instruments in a direction that explores policy questions and estimates the parameters of interest. Furthermore, MTE catches whole distributions of treatment effects, which makes it a useful tool for generating different treatment effects and comparing those effects. Using local IV estimation, Heckman and Vytlacil (2005) suggested a method to identify MTE as the derivative of the conditional expectation of the dependent variable with respect to the resistance to be treated (Cornelissen et al., 2016; Andresen, 2018). I use this method with the help of the Stata package *mtefe* to estimate peer effects in terms of MTE.

Since MTE can be estimated with a binary treatment variable, I generated a dummy variable indicating a high village average likelihood of HIV infection. A high subjective likelihood is likely to represent those who did not obtain test results because learning about these results is correlated with lower subjective likelihood. Focusing on high village subjective likelihood allows the MTE results to be comparable to IV estimates from the model 3.1. The dummy variable takes one if the village leave-out mean of subjective belief is larger than the median (0.0943 for current infection and 0.0879 for future infection). Each individual may have a different dummy value because the individual's personal belief determines the value. Therefore, different values might be assigned to those in the same village. Fifty per cent of the main sample takes the value of one for the dummy for the subjective likelihood of current infection. Furthermore, 47% of those in the main sample have a village mean subjective likelihood of future infection that is larger than the median.

Prior to the estimation of MTE, I estimated the effects of the high subjective likelihood of HIV infection on an individual's personal belief, using the following model:

$$Y_{ij} = \alpha + DY_{ij}\beta + X_{ij}\gamma + e_{ij} \tag{3.3}$$

where Y_{ij} is the subjective likelihood of HIV infection of individual i in village j, DY_{ij} is the dummy indicating the subjective likelihood of village neighbours larger than the median, and X_{ij} is the vector of controls, which are the same as those used in the model 3.1.

Table 3.9 presenting the estimation results, shows positive OLS and IV estimates for both the subjective likelihood of current and future infection. Living with village neighbours whose subjective likelihood of HIV infection is higher than the median likelihood is correlated with the individual's subjective likelihood of own HIV infection.

Table 3.9: Peer effects on subjective expectations with a dummy of low village likelihood

	Subjective likelihood			
	Infected now		Future infection	
	OLS	IV	OLS	IV
	(1)	(2)	(3)	(4)
High village mean likelihood	0.04***	0.09*	0.03***	0.12**
	(0.01)	(0.05)	(0.01)	(0.05)
N	1448	1448	1346	1346
F statistics (First stage)		12.2		11.1

*, **, *** indicate significance at the 10, 5 and 1%.
Covariates: age, age-squared, gender, education, owns land, married, district, distance to testing centre, distance > 1.5km, distance to locations in VCT zone, religion, subjective health, HIV status, and village means of covariates.

Equation 3.4 shows that the marginal treatment effect can be presented as the derivative of the conditional expectation of the dependent variable regarding the resistance to be treated, using the local instrumental variable (Heckman and Vytlacil, 2005). Y is the outcome and X presents observable characteristics. P(Z) shows the probability of entering treatment based on instrument Z. Therefore, $MTE(x, p)$ presents a marginal treatment effect with respect to the probability to be treated p, conditional on characteristics x.

$$MTE(x, p) = \partial E(Y \mid X = x, P(Z) = p)/\partial p \tag{3.4}$$

Regarding the effects on the subjective likelihood of HIV infection, the outcome variable Y indicates the subjective likelihood of HIV infection and X shows the control variables. Treatment is to be in a village where neighbours have subjective likelihood of HIV infection higher than the median. $P(Z)$ presents the probability of village neighbours having high subjective expectations, conditional on Z, monetary incentives to encourage them to obtain their test results. Treatment in this analysis is neighbours with high subjective likelihood of HIV infection. Since learning test results have negative effects on subjective likelihood of HIV infection (Table 3.5), treatment effect implies the effect of neighbours with high subjective HIV expectation who are unlikely to have learned their test results. The Marginal Treatment Effects disentangle the effect of peer's beliefs based on the identification assumption that one's incentives to learn test results, instrumental variables in this analysis, do not directly influence peers' HIV expectations.

Table 3.10 presents treatment effect parameters generated by the weighted average of marginal treatment effects. The treatment effect on the treated (TT), the effects of neighbours with high subjective expectations, is not significant. However, the treatment effect on the untreated (TUT), presenting the effects of village neighbours who learned about test results and have a low subjective likelihood of HIV infection, are positive and significant. The effect is larger than the average peer effect and the effect on those with a high subjective likelihood of HIV infection.[3] The LATE is positive and similar to the IV estimates in Table 3.9. Depending on the instruments, these results can explain the heterogeneity of peer effects among individuals.

MTE parameters show that peer effects are heterogenous among individuals. Peer effects are higher for those with village neighbours who learned their test results compared to those who did not. Being aware of one's HIV status is not only important for correcting subjective beliefs but also critical for affecting others' subjective expectations within the network. Therefore, encouraging individuals to obtain their test results can be a useful tool to facilitate peer effects on subjective expectations.

[3] The results are robust using a dummy for village neighbours with a subjective likelihood of infection larger than the mean instead of the median. See Appendix 3A.

Table 3.10: Marginal treatment effects parameters of peer effects on subjective expectations

	Infected now	Future infection
Average treatment effect (ATE)	0.14^{**}	0.18^{***}
	(0.06)	(0.06)
Treatment effect on the treated (TT)	-0.08	0.07
	(0.15)	(0.14)
Treatment effect on the untreated (TUT)	0.36^{**}	0.29^{*}
	(0.14)	(0.15)
Local average treatment effect (LATE)	0.12^{*}	0.17^{***}
	(0.06)	(0.06)

*, **, *** indicate significance at the 10, 5 and 1%. Treatment effect on the treated (TT) represents the effect of those with subjective expectations larger than median.
Treatment effect on the untreated (TUT) represents the effect of those who learned test results and have subjective expectations lower than median.

3.6 Robustness

Misreporting Expectations

The subjective likelihood of HIV infection data includes numerical measures with visual and interactive means. However, there is a possibility of misreporting subjective expectations because expectations are subjective, and HIV is a sensitive issue, especially in developing countries. Another possible reason for misreporting is that respondents may not understand the concept or method for measuring subjective expectations. For example, respondents might misunderstand the degree expressed by numbers from zero to ten for the subjective likelihood of HIV infection. Also, discerning between the likelihood of current and future infection can be confusing.

Table 3.11 presents the matrix by the number of responses classified by the subjective likelihood of current infection and the subjective likelihood of future infection. The same level of likelihood for both current and future infection is a possible response. A larger subjective likelihood of future infection than that of current infection is a possible response pattern for those who are more worried about their future HIV status than current HIV infection. However, the subjective likelihood of current infection being higher than that of future

Table 3.11: Subjective expectations of HIV infection comparing current and future infection

	Future infection											
Current Infection	0	1	2	3	4	5	6	7	8	9	10	Total
0	862	20	7	3	1	3	1	0	1	0	2	900
1	8	112	15	2	1	4	0	0	0	0	0	142
2	6	7	60	10	4	5	0	0	0	0	0	92
3	0	1	11	32	9	5	0	0	0	0	0	58
4	0	0	2	4	19	5	1	1	0	0	0	32
5	1	0	1	1	2	53	7	2	2	0	1	70
6	0	0	0	0	0	1	8	1	0	0	1	11
7	0	0	0	0	0	0	1	6	1	1	0	9
8	0	0	1	2	0	1	0	0	3	1	1	9
9	1	0	0	0	0	0	0	0	1	7	0	9
10	1	0	2	0	0	1	0	0	0	1	9	14
Total	879	140	99	54	36	78	18	10	8	10	14	1346

infection is not reasonable since the future status is less obvious, and there is no cure for HIV. These inconsistent responses might be due to misreporting caused by misunderstanding HIV, time differences or questions. Simple measurement errors or miscommunication during the survey might be another explanation for these responses.

Table 3.12: Subjective likelihood of HIV infection with corrected misreports

	N	Mean	SD
Current infection	1448	0.11	0.20
(Replaced with future infection)	1448	0.10	0.19
Future infection	1346	0.11	0.21
(Replaced with current infection)	1346	0.12	0.22

After revising these inconsistent responses, I repeat the regressions estimating peer effects on subjective expectations of HIV infection. In this regard, 66 responses with a larger subjective likelihood of current infection than future infection are subject to adjustments. Adjustments are made by replacing the subjective likelihood of current infection with future infection if the subjective expectation of current infection is larger than that of future infection. Another revision is the subjective likelihood of future infection being replaced by the subjective likelihood of current infection for those with a subjective likelihood of future infection smaller than that of current infection. After the revision, the average subjective likelihood of HIV infection is not substantially different from the initial subjective likelihood. However, the average subjective likelihood of future infection is slightly lower than the average likelihood of current infection. Therefore, the mean subjective likelihood of current infection partly replaced with the subjective likelihood of future infection is lower than before the adjustments. On the other hand, the average subjective likelihood of current infection partially replaced with the subjective likelihood of current infection is higher than before the adjustments.

I repeat the regression using the model 3.1 with the same instrumental variables. Table 3.13 presents the OLS and IV estimates for the peer effects within villages. Peer effects are positive and significant for both subjective expectations of current infection and future infection. The similar estimates and the significance of the modified estimates, considering possible misreporting, indicates the robustness of the estimated peer effects.

Table 3.13: Peer effects on subjective likelihood of HIV infection with corrected misreports

	Current infection		Future infection	
	OLS	IV	OLS	IV
Village mean	0.21***	0.57*	0.21***	0.63**
	(0.08)	(0.33)	(0.08)	(0.29)
N	1448	1448	1346	1346
F (First stage)		12.97		17.10

*, **, *** indicate significance at the 10, 5 and 1%.
Covariates: age, age-squared, gender, education, owns land, married, district, religion, distance to testing centre, distance > 1.5km, distance to locations in VCT zone, subjective health, HIV status, and village means of covariates.

Furthermore, I drop individuals who expect to have HIV now but not in the future and repeat the regression. Table 3.14 presents the OLS and IV estimates for the peer effects within villages. The results are robust after dropping observations with misreporting.

Table 3.14: Peer effects on subjective likelihood of HIV infection with dropped samples with misreports

	Current infection		Future infection	
	OLS	IV	OLS	IV
Village mean	0.21***	0.63**	0.23***	0.71***
	(0.07)	(0.32)	(0.08)	(0.27)
N	1391	1391	1289	1289
F (First stage)		12.18		19.92

Individuals who expect to have HIV now but not in the future are dropped. *, **, *** indicate significance at the 10, 5 and 1%. Covariates: age, age-squared, gender, education, owns land, married, district, religion, distance to testing centre, distance > 1.5km, distance to locations in VCT zone, subjective health, HIV status, and village means of covariates.

Number of Village Neighbours

The number of village members in the main sample varies largely from two to 85. Therefore, peer effects might only exist for those with few village neighbours in the main sample. Therefore, I repeat the estimation of peer effects without the villages with a small number of individuals in the main sample.

Table 3.15 presents the subjective likelihood of current HIV infection and future HIV infection for those without the small number of village neighbours in the main sample. 'Village Neighbours > 1' indicates those with at least two members in the same village included in the main sample. These villages include at least three individuals in the main sample. Samples are dropped gradually by the number of village neighbours up to three. For each exclusion, around ten to 30 individuals are dropped. Excluding individuals whose neighbours number up to three does not induce a remarkable change in the mean subjective likelihood of HIV infection.

I repeat the estimation of peer effects with reduced samples using model 3.1. Table 3.16 shows the peer effects estimates by IV regressions with different samples, including villages with different numbers of village neighbours. Peer effects for those with at least two village neighbours are positive and significant. Moreover, the effects are larger than peer effects

Table 3.15: Subjective likelihood of HIV infection with samples with different number of village neighbours

	Current infection			Future infection		
	N	Mean	SD	N	Mean	SD
Main sample	1448	0.11	0.20	1346	0.11	0.21
Village neighbours > 1	1430	0.11	0.20	1331	0.11	0.21
Village neighbours > 2	1412	0.11	0.20	1316	0.11	0.20
Village neighbours > 3	1384	0.11	0.20	1290	0.11	0.20

with the whole sample. Thus, the number of village neighbours in the main sample is not a substantial problem that changes the results.

Table 3.16: Peer effects on subjective likelihood with different number of village neighbours

	Current infection				Future infection			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Neighbours	> 0	> 1	> 2	> 3	> 0	> 1	> 2	> 3
Village mean	0.59*	0.70**	0.76**	0.78**	0.65**	0.95***	0.96***	0.95***
	(0.35)	(0.33)	(0.31)	(0.31)	(0.27)	(0.29)	(0.26)	(0.25)
N	1448	1430	1412	1384	1346	1331	1316	1290
F (First stage)	12.07	14.33	18.06	18.27	18.91	19.91	26.59	29.04

IV estimates. *, **, *** indicate significance at the 10, 5 and 1%.
Covariates: age, age-squared, gender, education, owns land, married, district, religion, distance to testing centre, distance > 1.5km, distance to locations in VCT zone, subjective health, HIV status, and village means of covariates.

3.7 Conclusion

Studying the formation and the effect of subjective expectations is important for health intervention. Peer effects might be useful to deal with HIV prevention in developing countries where information provision and social learning can play crucial roles. Learning the beliefs of network partners can provide information on possible HIV prevalence in an individual's environment and even warn of the severity of HIV. Understanding peer effects on subjective expectations supplement studies on social learning by investigating the effects of social networks on subjective belief. I study the peer effects on subjective expectations of an individual's HIV status in rural Malawi, and the data includes extensive subjective expectation data collected by visual and easy methods for the respondents. I estimate the peer effects of village neighbours on individuals' subjective likelihood of HIV infection influenced by HIV testing. Randomised monetary incentives to encourage them to learn about their test results are used as the source of variation to identify peer effects.

I find significant and positive peer effects of village neighbours on individuals' subjective likelihood of current and future HIV infection. IV estimates larger than OLS estimates for the peer effects might be due to measurement errors in the OLS estimates and the feature of IV estimates in which LATE are estimated. Peer effects on compliers whose neighbours are influenced by incentives to learn test results are larger than the effects on others. The estimated peer effects reflect the conformation of village neighbours, which is also evidence that perceived health risks depend on the common perceived risk within a social network. MTEs show the heterogeneity of peer effects depending upon the subjective expectations of neighbours. Significant effect on those whose neighbours with a low subjective likelihood of HIV infection implies those who learned about their test results have larger peer effects than others. Significant effects of village neighbours who are likely to have obtained test results show the possible role of HIV testing in forming subjective beliefs through social networks.

Subjective expectations can change health behaviour to prevent HIV infection regardless of actual HIV status. Peer effects on subjective expectations can prove an effective mechanism for providing information on HIV prevention in developing countries. Furthermore, beliefs can be affected not solely by the behaviours of network partners but by the subjective beliefs of neighbours. Significant peer effects in villages in rural Malawi show the possible role of village neighbours on subjective expectations regarding HIV. Using peer effects might be an effective method to correct subjective expectations and provide relevant information on HIV risk and prevention in sub-Saharan Africa. Larger peer effects for those who learned about their HIV test results with low subjective expectations of HIV risk emphasise the need to promote HIV testing.

Appendix 3A: Robustness checks for marginal treatment effect

As a robustness check for estimating marginal treatment effects, I repeat the estimation of MTEs using a dummy for village neighbours with a subjective likelihood of infection that is larger than the mean instead of the median. The average subjective likelihood of village neighbours is 0.1068 for current infection and 0.1026 for future infection. Fifty-seven per cent of the main sample take the value of one for current infection, and 55% take the value of one for future infection. More respondents are around neighbours with a subjective likelihood that is larger than the mean compared to those with neighbours with a subjective likelihood that is larger than the median.

Firstly, I estimated peer effects by regressing one's own belief of HIV infection on the dummy, indicating a high likelihood of infection for village neighbours with an OLS and IV using randomised monetary incentives as instrumental variables. The results show positive and significant estimates.

Table 3.17: Peer effects on subjective expectations

	Subjective likelihood			
	Infected now		Future infection	
	OLS	IV	OLS	IV
	(1)	(2)	(3)	(4)
High village mean likelihood	0.03**	0.09*	0.04***	0.14***
	(0.01)	(0.05)	(0.01)	(0.05)
N	1448	1448	1346	1346
F statistics (First stage)		10.4		13.9

*, **, *** indicate significance at the 10, 5 and 1%.
Covariates: age, age-squared, gender, education, owns land, married, district, distance to testing centre, distance > 1.5km, distance to locations in VCT zone, religion, subjective health, HIV status, and village means of covariates.

The MTE parameters are similar in size and significance compared to the effect estimated with treatment to be in village with a subjective likelihood higher than the median, except for a treatment effect on the untreated for the likelihood of current infection. LATE are significant and similar to IV estimates. This check confirms that the marginal treatment effects are robust, and the results are not from the treatment variable creation process.

Table 3.18: Marginal treatment effects parameters

	Infected now	Future infection
Average treatment effect (ATE)	0.16**	0.26***
	(0.07)	(0.06)
Treatment effect on the treated (TT)	0.17	-0.04
	(0.15)	(0.13)
Treatment effect on the untreated (TUT)	0.16	0.46***
	(0.16)	(0.14)
Local average treatment effect (LATE)	0.15**	0.21***
	(0.07)	(0.05)

*, **, *** indicate significance at the 10, 5 and 1%.

Chapter 4

Extended family networks and rainfall shock

4.1 Introduction

Individuals in the rural areas of developing countries face various types of risks and uncertainties. Within-village extended family networks, and the characteristics of interactions in these networks, play an especially important role in risk and resource sharing. Imperfect credit markets and insufficient social security arrangements lead households to depend on informal risk-sharing among extended family networks in developing countries (Angelucci et al., 2017). Extended families are also tightly connected, making social enforcement within networks as an advantage for informal insurance (Ferrara, 2003). However, little is known about how the structures of these networks are formed and affected by income shocks.

Excessive or insufficient amount of rainfall creates risks for agricultural income in developing countries. Rainfall shocks are known to have significant impacts on the important decisions of families in developing countries, such as education, employment, early marriage and migration (Munshi, 2002; Corno et al., 2020; Adhvaryu et al., 2021). In terms of marriage, the decision to marry and the marriage location choices can be affected by rainfall events (Rosenzweig and Stark, 1989). Furthermore, living arrangements are mentioned to provide inter-generational support and facilitate the sharing of public goods (Rosenzweig and Wolpin, 1993; Salcedo et al., 2012). Thus, changes in living arrangements to co-reside or split households are likely to be motivated by economic reasons and affect risky decisions made by households. Therefore, studying the responses of extended family networks within villages to rainfall shocks would provide interesting insights into the role of extended family networks in risk sharing. Although existing literature has shown how rainfall shocks affect individual decisions that shape the formation of these networks (marriage or migration), there is no evidence of existing literature that has studied the overall effect, which could have different directions.

This chapter studies the influence of rainfall shocks – measured as deviations from long-run average rainfall – on the structure of within-village extended family networks in rural Mexico. The study explores the overall effect of rainfall shocks on extended family networks. Furthermore, the effects of conditional cash transfers will be analysed. The mitigation of liquidity constraints caused by conditional cash transfers may change the effect of income shocks on within-village extended family networks. Although studies emphasised the importance of within-village extended family network, not many studies analysed the network statistics and the changes on the network statistics of within-village extended family network. Furthermore, this chapter can contribute to the literature by estimating the overall effect of rainfall shocks on possible mechanisms of within-village extended family network including marriage, migration, and split households.

To study the characteristics of extended family networks within villages, we focus on extended family network data created from a household survey performed to evaluate the PROGRESA programme in rural Mexico. PROGRESA was launched by the Mexican government to enhance the health, educational and nutritional status of low-income families in marginalised villages. Conditional cash transfers are assigned to eligible (poor) households. The conditions were enrolling and attending school and visiting health care centres regularly. For the assessment of PROGRESA, a randomised controlled trial (RCT) was conducted. The experimental design of the RCT assign village into treatment group and control groups. This random assignment provides the source of variation for the evaluation of the programme.

Data from the panel survey contain very comprehensive information, and the assessment involves all households in selected villages. This feature relieves concerns about data availability, which is a usual issue in studies on extended family networks. Using the data from the poverty-alleviation programme can be a meaningful approach in that the analysis can also reflect how social programmes or policies can change the economic decisions of households in developing countries. This approach extends the analysis to the effects of policy interventions on the network formation process, which can either counteract the effect of rainfall shock or reinforce the influence.

Within-village extended family networks are created from this data by discovering links between households using full name information. The algorithm is inspired by the tradition in Mexico to grant individuals with two surnames, one from the father's lineage and the other from the mother's lineage. The complete details on extended family networks generated from all households within villages enable a thorough analysis of the effect of rainfall shocks on network structures.

We use rainfall data matched at the village level from the closest rain gauge to the village. Rainfall shocks are measured as an absolute deviation of rainfall from the long-run average to consider the recorded level of rainfall compared to the normal amount. We also define rainfall shock as a dummy equal to one if the absolute deviation of rainfall is one standard deviation above or below the long-run average and the absolute deviation from the long-run average divided by the standard deviation. The exogenous feature of rainfall shock allows the regressions of network characteristics on rainfall shock variables to disentangle the effect of the shock from other influences of unobserved variables.

Our findings show that in villages where a reduction in agricultural incomes is implied due to rainfall shocks, extended family networks have a smaller network degree and size. However, conditional cash transfer from PROGRESA counteracted this negative effect. Extended family network structures can be changed due to the decisions of households to split, marry or migrate. Generating a separate household after marrying within the village and leaving one's household can increase the extended family network degree and size. In addition, when entire households migrate out of a village for employment, the degree and size of the network can be reduced. The decisions might be influenced by the speed of recovery from rainfall shocks or the mitigation of liquidity constraints from cash transfers. For example, some households may delay splitting households since they might not have enough financial resources to pay for another set of household-related public goods for the new family unit. Therefore, we examine the effect of rainfall shocks and cash transfers on marriage and migration to explore the drivers of changes in the extended family network.

Household members who experienced rainfall shocks are less likely to leave their households for marriage. This finding suggests that rainfall shocks might be associated with households' decisions not to split, causing smaller degrees and sizes of extended family networks. Work migration reduced by rainfall shock in villages with cash transfers shows the effect of financial transfers in countervailing the negative effect of rainfall shocks in terms of network degree and size. Cash transfers are possible means to mitigate the effect of rainfall shocks, serving to counteract the reduction of extended family networks degree and size in villages by alleviating liquidity constraints. Therefore, the creation and transformation of extended family networks within villages are important processes warranting consideration when studying the impact of income variations. Households can change their living arrangements after experiencing rainfall shocks as a means through which the effects on income could be addressed. Furthermore, PROGRESA changing the effect of rainfall shocks indicates that relevant policies can alleviate the liquidity constraints of those affected in developing countries, even with income variations.

4.2 Literature

An extended family network is considered a meaningful institution through which economic functions are performed in developing countries. Social enforcement under the norm of collective responsibility within a kin group is a feature different from normal communities. Furthermore, reciprocation can be performed well with the help of family members. Ferrara (2003) claimed these two advantages as the basis of the extended family network having superiority over informal insurance. The nature of the extended family network, where preferences and inter-generational investments exist in extended families, further assists the importance of the extended family network (Angelucci et al., 2009). These networks are known as being capable of achieving the motivation from altruism and the power to impose obligations (Angelucci et al., 2017). Indeed, substantial amounts of financial transactions are conducted within extended family networks, including loans and transfers, exceeding the scale of transactions in nuclear families (Lucas and Sark, 1985).

Within-village extended family networks can be vital institutions for risk and resource sharing in developing countries. Angelucci et al. (2009) mentioned that the meaningful functions of extended family networks are consumption smoothing and investments for wealth status in the village economy. In addition, the high costs of forming inter-village networks are fundamental issues relevant to the value of intra-village extended family networks. Using PROGRESA data from rural Mexico, Angelucci et al. (2017) showed that extended family networks could pool resources to deal with liquidity constraints. The study insisted that cash transfers from PROGRESA generated investments for households connected to extended family networks rather than for isolated ones. Informal risk sharing with insurance within a village played a significant role for households suffering from idiosyncratic shocks in consumption (Kinnan et al., 2020). The enforcement of informal arrangements was meaningful in the role of risk-sharing of socially close connections, such as directed connections within the family network. Furthermore, socially distant connections, including indirect links in extended family networks, could be valuable for risk-sharing, the changes of which are critical (Malde, 2016). A paper evaluating the effect of cash transfers on school enrolment provided suggestive evidence that resources are shared in extended family networks, from households that receive enough cash transfers for enroling their children in primary school to those receiving less attention for secondary school enrolment for their children (Angelucci et al., 2009). Kinship networks could supplement financial networks in developing economies by allowing transactions that are considered too large to be collateralised with tangible assets (Kinnan et al., 2012). Although several studies mentioned extended family networks as an important system in developing countries, not many studies investigated the formation and structure of extended family networks.

Marriage, migration or other household arrangements are suggested as ways to address adverse income shocks in developing countries. Extended family networks within villages can be created and changed in response to these activities. Rosenzweig and Stark (1989) argued that marital arrangements play substantial roles in facilitating smooth consumption and overcoming income risks. Furthermore, households in rural India with more changeable agricultural incomes are reported as having a higher likelihood of risk pooling via marriage migration. Corno and Voena (2016) estimated the effect of income shocks on child marriage in Tanzania. Families with teenage girls who experienced rainfall shock are likely to marry off the girls by the age of 18. Another study by Corno et al. (2020) examined the effect of droughts on the timing of marriage for women in sub-Saharan Africa and India, where the direction of marriage costs differs. The study found positive effects of rainfall shocks on early marriage in sub-Saharan Africa, where bride price payments are prevalent. However, in India, where dowry payments are customary, droughts decreased the hazard of early marriage.

Studies have found that abnormal amounts of rainfall increase migration in regions where agricultural outcome matters for developing country economies (Henry et al., 2004; Puente et al., 2015). These findings infer that migration induced by income variations due to weather is another important factor that affects extended family networks. Using migration data in rural Mexico, a study found that a migrant household is likely to receive new members around migration episodes. The study argued that the high probability of migrant households generating the attrition in the study of migration suggests that household members of migrants may have disbanded their households and joined other households (Bertoli and Murard, 2017). Although migration is a common mechanism to avoid income variations in agrarian societies, credit constraints and risk aversion might restrict such movement (Bryan et al., 2014).

Adult children co-residing with their parents is another suggested living arrangement facilitating inter-generational support since shared residence and inter-household transfers are important methods to allocate resources (Rosenzweig and Wolpin, 1993). Kaplan (2012) pointed out labour market events as possible causes for the timing of young children to co-reside with their parents or leave their parents' households. In addition, household size or arrangements might be linked with income because a decrease in the consumption of public goods, implied by higher income, would lead to smaller households with a diminished need to share the costs of public goods among household members (Salcedo et al., 2012). We supplement this literature by studying the possible drivers of the effect of income risks on within-village extended family networks, such as marriage, migration and inter-generational co-residence. Moreover, the literature has studied the impacts of rainfall shocks on each of these decisions (marriage, migration and co-residence) separately. Thus, we explore the effects of rainfall shocks on all decisions altogether and study the effects on the overall network.

In addition, studies evaluated the effect of policies or programmes on the formation and change of networks in developing countries. Hess et al. (2020) analysed a large Community-Driven Development (CDD) program in The Gambia, which provided funding for households in 500 randomly selected villages. The program allocated enough funding, approximately half the country's GDP per capita, and the treatment reduced informal economic transactions. This result is mentioned as being driven by a modest increase in economic well-being and unequally distributed benefits. Banerjee et al. (2021) found that the introduction of microfinance diminished social networks in rural villages in India. This study used detailed information on social networks and found an interesting spillover effect on reducing social networks in those with a low likelihood of joining microfinance. These studies suggest that policies or programmes established for economic development in developing countries may have consequences on networks, such as fewer social connections and network spillovers.

The effects of PROGRESA – a poverty-mitigation programme in rural Mexico that provides cash transfers – on family or living arrangements are assessed by a couple of studies. Bobonis (2011) found that families eligible for PROGRESA increased marital dissolution rates, although no significant change occurred in the overall share of women in a marital union. Conditional cash transfer from PROGRESA also increased international migration for labour, whereas secondary school subsidies reduced short-term migration (Angelucci, 2004). We supplement the literature by exploring the influence of PROGRESA on network characteristics with the creation of within-village extended family networks.

4.3 Data

4.3.1 Background of the programme: PROGRESA

The analysis uses household survey data collected to evaluate the previously mentioned social assistance programme called PROGRESA, conducted in rural Mexico. PROGRESA was launched in 1997 by the Mexican government to enhance the health, educational and nutritional status of low-income families, especially for children and mothers of low-income households. Conditional cash transfers are provided to the mothers of targeted families. The conditions are enrolling and attending school and visiting health care centres regularly. PROGRESA reached around 2.6 million families, representing one-ninth of all families in Mexico at the end of 1999. Not only is the enormous amount of coverage an advantage of this programme, but this programme differs from earlier social programmes in that it directly targets extremely low-income families, focusing on multiple dimensions of human capital (Skoufias, 2005).

Household panel data were collected from 24,000 households in 506 villages every six months from October 1997 to November 1999. Furthermore, 320 villages were randomly assigned to the treatment group that was implementing PROGRESA among 506 villages in rural Mexico in which the programme was assessed. The survey round in October 1998 was the first wave after the beginning of the cash transfers. Households were classified as eligible or ineligible in determining the beneficiaries of the cash transfers. Eligibility (low-income households) was determined by a household welfare index: a weighted average of (a) household income; (b) the number of household members; (c) ownership of durables, land, and livestock; (d) education; and (e) physical features of the dwelling (Angelucci et al., 2006). Cash transfers were provided every two months, and the average monthly amount accounted for 20% of monthly consumption expenditures.

4.3.2 Survey data

Data from the October 1998 survey is used for analysis because it is the first survey wave started in the post-programme period, during which the names of respondents were collected. The information on names is essential for deriving extended family networks from the data. Two features make the data suitable for the analysis. Firstly, the survey data provides exceptionally exhaustive information since data was collected from all households in 506 rural villages. The survey collected household rosters of all households and comprehensive socio-economic information, such as consumption, income, household demographics and migration. The second feature of the data is that the programme is designed to randomly assign villages into treatment and control groups. These two advantages offer a flexible approach for evaluation by allowing any type of estimator to assess the programme's impact (Skoufias, 2005).

4.3.3 Within-village extended family network

Within-village extended family network data are generated by modifying the algorithm in Angelucci et al. (2009), connecting households with surnames. This mapping algorithm is possible due to the convention of Mexico assigning individuals two surnames – a paternal and maternal surname. The panel survey data collected complete details on the names of respondents – (i) paternal surname, (ii) maternal surname and (iii) given name. Four surnames are available from each household, namely, paternal and maternal surnames of household heads and spouses. Then, we match these names and identify links between households. Extended family networks are formed via two types of family links. One type is the intra-generational link, including siblings of the household head or household spouse. Another type is the inter-generational family link, including connections from the household head and household spouse to their adult sons or daughters and links from the household head or household spouse to their parents.

Firstly, intra-generational links can be identified for those (household head or spouse in each household) with the same maternal and paternal surnames. Next, inter-generational links can be identified because children adopt paternal surnames from both their father and mother. Thus, we define an inter-generational link between household 'A' and household 'B' if: (a) the head or spouse of 'A' has the paternal surname the same as the paternal surname of the head of 'B' and (b) the maternal surname of the head or spouse of 'A' is identical with the paternal surname of the spouse of 'B'. We define extended family networks by identifying all possible intra-generational and inter-generational links among households within the village.

Lastly, age restrictions are imposed to eliminate spurious family links. The age difference between the oldest and the youngest in each sibling group cannot exceed 30 years for sibling groups. For parent–child links, the mother needs to be at least 15 years older than her eldest child, and at most 45 years older than her youngest child.[1] As a result, 2,477 extended family

[1]To avoid spurious connections and the reduction of strength in the algorithm, one village with a large

networks within villages are generated, including 16,399 households.[2]

Network characteristics of interest are degree, clustering coefficient, size and average path length (the definitions of network characteristics are introduced in Appendix 4A). These indicators are deemed fundamental statistics that present meaningful features of network structures (Jackson, 2008). Notably, 16,399 households in the main sample belong to 2,477 within-village extended family networks. Degree and clustering coefficients can be calculated at the household level. The size, average path length and clustering coefficient are calculated at the network level.

Figure 4.1 presents a graphical representation of extended family networks within two villages as examples of network structure. Two villages have differences in terms of network characteristics because most networks, except isolated households in Village B, are fully connected networks where all households are interconnected. Thus, Village B's degree and clustering coefficient are larger than Village A's. On the other hand, the average path length is larger in Village A due to more pairs of households that are not directly linked. Due to one large network connecting several households to the bottom right corner of Figure 4.1a, the average size of the network is larger in Village A than Village B. Network statistics can reflect different aspects of network characteristics depending on the indicator.

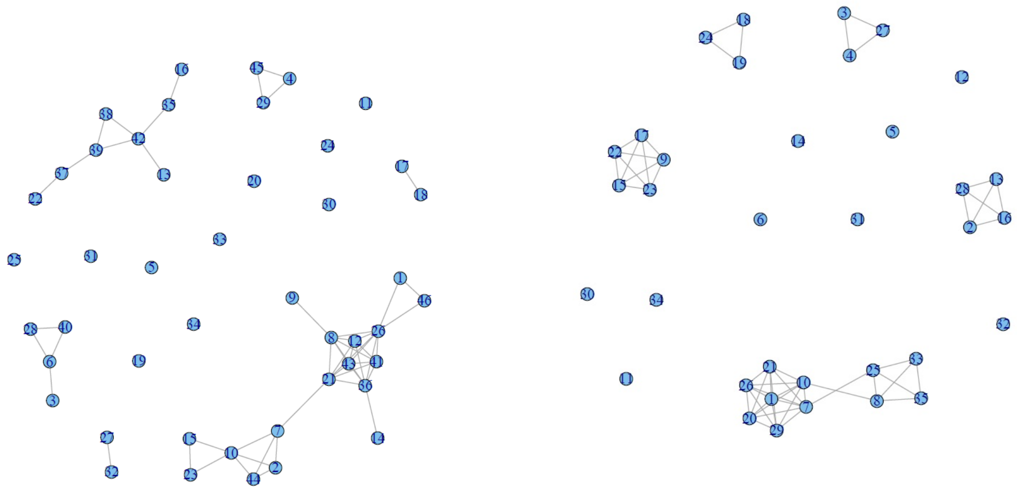
Table 4.1: Within-village extended family network statistics

	N	Mean	SD
Household level			
Degree	16399	4.11	4.39
Clustering coefficient	12339	0.71	0.33
Network level			
Clustering coefficient	1413	0.58	0.34
Size	2477	7.03	13.46
Size / village size	2477	0.13	0.17
Average path length	2477	1.44	0.79

Network statistics are described in Appendix 4A.

portion of individuals with similar surname combinations is dropped. Furthermore, networks with sizes exceeding 100 households are dropped.

[2] Around 20% of observations were isolated households. The algorithm does not include these households in any within-village extended family network.

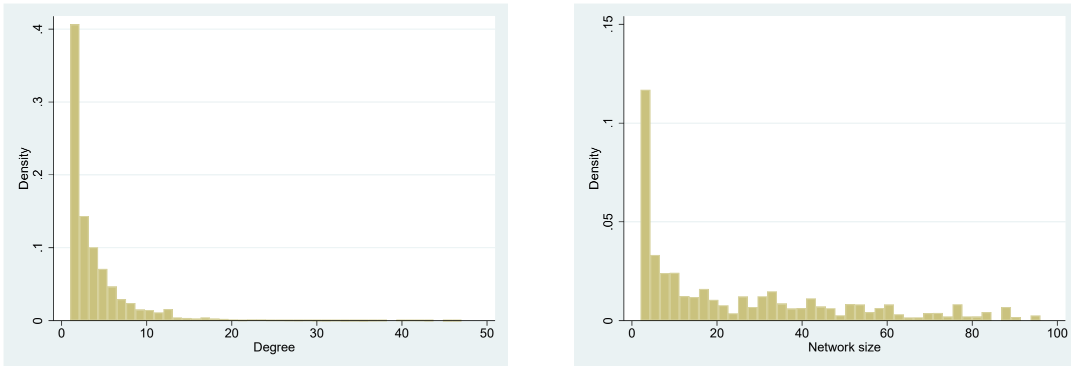


(a) Village A (b) Village B

Figure 4.1: Family network graphs

Table 4.1 presents the network characteristics of the within-village extended family network. The degree can represent the extent of the household's connectedness. Furthermore, each household is, on average, connected to four households. The clustering coefficient can essentially show the network's density at the local level (Jackson, 2005). The clustering coefficient is 0.71 at the household level and 0.58 at the network level on average. The sample does not include isolated households. Each extended family network, except those with a size of one, comprises an average of seven households, as shown by the size of the network. The size of the network divided by the village size (the number of households in the village) is 0.13. This measure implies the importance of family networks with a size larger than one relative to the size of the village (Angelucci, 2009). The average path length is often used to indicate network efficiency, showing how the transportation of information flow could be conducted through the network (Ye, 2010). The average path length of a within-village extended family network is 1.44 on average.

The figures 4.2 present the distributions for the degree and size of an extended family network within a village. Network degrees are concentrated at the value of one, implying that a large portion of extended family networks are dyads. The degree of most households



(a) Degree (b) Size

Figure 4.2: Degree and size distribution

is smaller than ten, implying that there are not too many spurious connections. Regarding size, the largest portion of extended family networks comprises two households.

Examining the effect on degree could show the change in the connectedness of within-village extended family network. Rainfall shock may create new connections due to the need for risk sharing through networks. However, income shocks may reduce connections of extended family networks by combining households or disconnecting links. The changes in the size of network could also reflect these possible reactions including merge and disconnections among households to rainfall shocks. In addition, clustering coefficient may be correlated with better enforcement of informal arrangements since having common links means there could be more stringent punishment if one deviates from an informal arrangement. The impact on average path length may show the influence of rainfall shocks on the need for denser connections of extended family network within villages for risk sharing. Efficient flow through extended family networks could be generated after experiencing difficulties after rainfall shocks.

4.3.4 Rainfall data

Rainfall data is collected from the Mexican Weather Service to examine the effect of rainfall shocks. The data recorded actual rainfall from January 1985 to January 2005. Thus, rainfall data from 1985 is used to generate long-run average rainfall. Robustness checks are conducted using the rainfall from 1990. Rainfall data matched at the village level in the main sample exploited the closest rain gauge to the village. The average distance to the nearest rainfall gauge was 18km. Monthly rainfall amounts were aggregated yearly. Thus, monthly rainfall from November 1997 to October 1998 (year t) was added since the survey data were collected in October 1998. Rainfall in year t-1 was the sum of monthly rainfall from November 1996 to October 1997. The average rainfall was around 1,244mm in 1998 and 1,052mm in 1997. The 15-year average and 15-year standard rainfall deviation were used as long-run rainfall indicators. Rainfall in 1998 was higher than the long-run average rainfall, which is likely to be caused by El Nino in 1997 and 1998, accompanied by the Pacific jet stream and increased rainfall. The coefficient of variation, calculated as the measure of riskiness (15-year average / 15-year standard deviation), was 0.3 on average.

Table 4.2: Rainfall statistics

	N	Mean	SD
Year t (1998)	16399	1243.70	742.40
Year t-1 (1997)	16399	1051.74	651.25
15-year average	16399	1113.64	655.55
15-year standard deviation	16399	304.89	179.63
Coefficient of variation	16399	0.30	0.15
Distance to the nearest rainfall gauge	16399	17.87	9.03

4.3.5 Main sample

From the October 1998 round in the PROGRESA household survey, 16,399 households belonging to extended family networks and without missing variables are selected as the main sample. Table 4.3 presents the descriptive statistics of the main sample. Most household heads were male, and 88% of them had work. Among the six household members, the number of male adults was around 1.5 on average.

Understanding the different reasons for the change in the composition of households is indispensable in determining what derives the extended family network structure within the

Table 4.3: Descriptive statistics

	N	Mean	SD
HH head age	16399	46.16	15.10
HH head male	16399	0.94	0.23
HH head has work	16399	0.88	0.32
HH head had education	16399	0.71	0.45
HH poor	16399	0.54	0.50
HH size	16399	5.93	2.71
Number of male 12-18	16399	0.49	0.76
Number of male adults	16399	1.46	0.85
Number of female 12-18	16399	0.47	0.75
Number of female adults	16399	1.44	0.80
Marriage migration out of village	16399	0.01	0.10
Work migration out of village	16399	0.04	0.19
Left HH within village	16399	0.02	0.13
Left HH within village for marriage	16399	0.01	0.10
Treatment (PROGRESA) village	16399	0.63	0.48
HH eligible for treatment	16399	0.54	0.50

village. Therefore, variables are generated to present whether individuals left their household separately by destination (out-of-village or within-village). The variables are generated for those aged 15–39 who left their households 12 months prior to the survey being conducted in October 1998. The variables indicate leaving 16,399 households in the sample either as migrating out of the village or moving around within the village. In this regard, 1% of the main sample migrated out of the village for marriage, and 4% migrated for labour. Anyone who left their households within the village can imply split households. Accordingly, 2% of households in the main sample had individuals who left households within the village, and 1% were experienced individuals leaving households for marriage within the village.

The experimental design of the PROGRESA programme with random allocation of treatment by villages allows for evaluating the programme with a relevant identification strategy using the variation of treatment. Sixty-three per cent of the main sample were randomly assigned to the treatment group for cash transfers. Households were classified as eligible (low-income) households and ineligible (not low-income) households for the treatment. In this regard, 54% of households were assigned to poor villages by criteria including income, the size of the household, ownership of durables, education and house conditions. Cash transfers are provided to these households allocated as eligible households.

4.4 Empirical Framework

4.4.1 Model

To study the effect of rainfall on within-village extended family network characteristics, we examine the following model:

$$N_{ijk} = \alpha + Rain_{ijk}\beta_1 + P_{ijk}\beta_2 + (Rain_{ijk} \times P_{ijk})\beta_3 + X_{ijk}\gamma + \theta_k + e_{ijk} \tag{4.1}$$

where N_{ijk} denotes network characteristics of household i in network j and state k, $Rain_{ijk}$ denotes rainfall shock, P_{ijk} presents PROGRESA exposure of household i in network j and state k, which is cash transfer provided as treatment, and $Rain_{ijk} \times P_{ijk}$ indicates the interaction term of rainfall shocks and treatment. β_1 represents the effect of rainfall shocks, β_2 shows the effect of PROGRESA for individuals who did not experience rainfall shocks, and β_3 indicates the effect of cash transfer for those who are in villages with rainfall shocks. In other words, β_1 is the effect of rainfall shocks in control villages, and β_3 is the effect of rainfall shocks in treatment villages. Standard errors are clustered at the village level.

Average path length and size can only be examined at the network level. For those characteristics, we estimated the same model as in Equation 4.1 at the network level with network mean variables.

4.4.2 Identification strategy

The analysis can examine how the effects of rainfall shocks on within-village extended family networks vary by treatment status of PROGRESA. The model specification takes advantage of the exogeneity of both rainfall shocks and PROGRESA exposures. The experimental design that randomly assigned treatment villages for cash transfer can be the source of variation. These exogenous variations provide advantages to imply the causal effect of rainfall shocks on within-village extended family networks. The main assumption for this model specification is that rainfall shocks are exogenous, so they are orthogonal to unobserved determinants of outcome variables. This assumption is rational because weather phenomena are exogenous and random. Therefore, we use OLS estimation to identify the effect of rainfall shocks based on treatment status, using the variation of rainfall shocks and treatment assignments.

Rainfall is often used as an exogenous variation for estimations related with income in agrarian economies since farmers rely on the timing and the amount of rainfall. In this context, rain-fed subsistence agriculture is a key economic activity. Rainfall could affect extended family network through income shock, whereas extended family network characteristics do not change the timing and the quantity of rainfall. The households in this context cannot do anything to influence rainfall. Thus, rainfall shock is not likely to be correlated with the residuals of our regressions estimating the influence of rainfall on within-village extended family network.

This analysis use cross-sectional variation. Using time variation with more surveys would broaden perspectives and deepen the implications of the study. As an initial analysis, current study examines the effect of rainfall shocks using the first survey after conditional cash transfer. Exploiting data from consecutive surveys will need substantial amount of work to generate extended family network data and calculate network characteristics. Future research with time variation from additional data would be meaningful and show dynamic effects which could be more complicated.

4.4.3 Rainfall shock variables

Rainfall shocks are measured as an absolute deviation of rainfall from the long-run average to consider the level of rainfall compared to the normal amount because the average amount of rainfall varies by village depending on terrain and location. Using the deviation of rainfall from the mean as a rainfall shock variable is a common way to identify weather shocks in 'levels' (Dell et al., 2014). Adhvaryu et al. (2018) emphasised the relevance of using this relative measure instead of the absolute amount of rainfall because the identical rainfall amount can have different implications for villages with different average amounts.

We also define rainfall shock as a dummy equal to one if the rainfall is one standard deviation above or below the long-run average. This definition focuses on villages that have experienced an unusual amount of rainfall. It is necessary to consider that a simple comparison between villages with a lot of rainfall and villages where the typical rainfall amount is minimal might result in inappropriate conclusions (Adhvaryu et al., 2018). Furthermore, the measure is useful for figuring out nonlinear effects by indicating the frequencies of rainfalls that fall into the categories of different amounts (Dell et al., 2014).

Lastly, absolute rainfall deviation from the long-run village mean divided by the long-run village standard deviation is selected as another measure of rainfall shock. This approach adopts the climate–economy model where level differences are important in proportion to an area's normal variation instead of the level itself (Dell et al., 2014).

Table 4.4 presents the means and standard deviations of three rainfall shock measures used in the analysis. Long-run average and long-run standard deviation are derived for 15 years: the longest time for available data. The absolute deviation is divided by 1000 to present rainfall amount in metres, and the absolute deviation is 190mm in year t on average. This measure is 0.68 on average if divided by the 15-year standard deviation of rainfall. 30% of the main sample experienced rainfall amount with an absolute deviation one standard deviation above or below the 15-year mean. The same rainfall shock variables are generated for year t-1 (November 1996 to October 1997). These lagged variables are included to examine the effect appearing with time differences. Households may spend time on decisions resulting in changes to the extended family network structure. Absolute deviations from the long-run village mean are similar in year t and year t-1, whereas the portion of the main sample with an absolute deviation one standard deviation above or below the 15-year average rainfall is larger in year t than that in year t-1. Since the first cash transfers were initiated in March 1998, the effect of rainfall shocks in t-1 for the treatment group will indicate how cash transfers have accelerated the recovery after rainfall shocks.

Table 4.4: Rainfall shocks

	N	Mean	SD
Absolute deviation from 15y avg / 1000 at t	16399	0.19	0.25
Absolute deviation from 15y avg / 1000 at t-1	16399	0.19	0.20
Absolute deviation 1sd above/below 15y avg at t	16399	0.30	0.46
Absolute deviation 1sd above/below 15y avg at t-1	16399	0.21	0.41
Absolute deviation from 15y avg / sd at t	16399	0.68	0.57
Absolute deviation from 15y avg / sd at t-1	16399	0.63	0.53

The analysis includes rainfall shock variable in year t and year t-1 in the same regression. Contemporaneous shocks and lagged shocks are in the same regression because the shocks can have different effects depending on the timing. For example, migrations occur not only permanently due to the incentives to search for well-paid long-term jobs, but also temporarily to cope with immediate shocks. Contemporaneous shocks are more related with temporary migration, such as moving to near villages and coming back later. Furthermore, marriage is costly and takes time to build saving. Thus, network changes because of marriage are likely to be related to lagged shocks. Lagged shocks can show the longer lasting effects and how networks recover from the impact of rainfall shocks. The contemporary shock implies an immediate income loss, while the lagged shocks capture longer term effects of the income losses, which may operate through reduced assets and wealth.

4.5 Results

4.5.1 Effect on network structure

OLS regressions are used to estimate the effects of rainfall shocks on within-village extended family network statistics. Regressions are conducted for each measure of rainfall shocks: (i) absolute deviation of rainfall from the 15-year average, (ii) a dummy indicating whether rainfall amount is one standard deviation above or below the 15-year average, and (iii) absolute deviation of rainfall from the 15-year mean divided by 15-year standard deviation. For each measure, the effects are estimated (a) for rainfall shock and interactions term in year t, and (b) with rainfall shock and interaction terms in both year t and year t-1.

Table 4.5: The effect of rainfall shock on degree

	Degree					
	(1)	(2)	(3)	(4)	(5)	(6)
Rainfall shock	Abs deviation		1sd above/below		Abs dev/sd	
Shock (t)	-1.82**	-1.84**	-0.78	-0.84	-0.72	-0.85
	(0.85)	(0.86)	(0.74)	(0.73)	(0.63)	(0.64)
Shock (t-1)		-1.24		-0.86		-0.87*
		(1.10)		(0.58)		(0.50)
Shock (t) * Treatment	1.13	0.91	-0.04	0.07	0.03	0.18
	(1.01)	(0.99)	(0.78)	(0.78)	(0.70)	(0.70)
Shock (t-1) * Treatment		3.28**		1.22*		1.34**
		(1.28)		(0.69)		(0.54)
Treatment	-0.43	-0.99*	-0.22	-0.50	-0.26	-1.21
	(0.46)	(0.58)	(0.49)	(0.58)	(0.66)	(0.92)
N	16399	16399	16399	16399	16399	16399
Average degree	4.11	4.11	4.11	4.11	4.11	4.11

OLS estimates. *, **, *** indicate significance at the 10, 5 and 1%. SE clustered by village. Column 1-2 represent the effect of rainfall shocks defined as absolute deviation from long-run average. Column 3-4 represent the effect of rainfall shocks defined as a dummy if absolute deviation is 1 standard deviation above or below long-run average. Column 5-6 represent the effect of absolute deviation of rainfall from long-run average divided by standard deviation. Covariates: HH head age, age-squared, gender, work, education, HH poor, size, composition, and state FE. "Treatment" denotes cash transfer from PROGRESA.

Table 4.5 presents the effect of rainfall shock on extended family network degree within villages. Columns 1-2, 3-4 and 5-6 use different definitions of rainfall shocks (see the table notes). Columns 1, 3 and 5 represent the effect of rainfall shocks in year t. Columns 2, 4 and 6 include the lagged term of rainfall shock with a lag of one year. These lagged terms are included to estimate the effect of time differences and whether cash transfers accelerated the recovery after the rainfall shock from the previous year. The treatment presents the effect of conditional cash transfer from PROGRESA.

The effect of rainfall shock on the degree of the extended family network in the control group is negative. Furthermore, only the negative effect of the absolute deviation from the long-run average rainfall is significant. The negative effect of the lagged rainfall shock is only significant for the absolute deviation divided by standard deviation. Interaction effects (Rainfall shock * Treatment) in lagged terms are positive and significant for all definitions of rainfall shock. Since lagged rainfall shocks in the control group have a negative effect, cash transfer countervails the effect of rainfall shock from year t-1. This finding implies that the treated households recover faster in terms of their degree.

Table 4.6 represents the effect of rainfall shock on extended family network size within villages denoted by the number of households. Columns 1-2, 3-4 and 5-6 use different definitions of rainfall shocks, and Columns 2, 4 and 6 include the lagged term of rainfall shock with a lag of one year. The effect of rainfall shock on the extended family network size is negative in the control group. The negative effect of rainfall shock is only significant for absolute deviation from the long-run average. The effects of lagged rainfall shocks are not significant. However, the effect of rainfall shocks in the treatment group is positive. Lagged interaction effects (Rainfall shock * Treatment) using absolute deviation from the long-run average and absolute deviation divided by standard deviation are significant. Cash transfers offset the negative effect of rainfall shock, implying the acceleration of recovery in terms of network size.

Table 4.7 represents the effect of rainfall shocks on average path length. The effect of rainfall shocks is positive in the control group and negative for lagged interaction terms in the treatment group. The effects are small and insignificant.

Table 4.6: The effect of rainfall shock on network size

	Network size					
	(1)	(2)	(3)	(4)	(5)	(6)
Rainfall shock	Abs deviation		1sd above/below		Abs dev/sd	
Shock (t)	-2.91**	-2.84**	-0.60	-0.65	-0.59	-0.70
	(1.35)	(1.35)	(0.86)	(0.85)	(0.75)	(0.75)
Shock(t-1)		-1.70		-0.88		-0.89
		(1.45)		(0.87)		(0.60)
Shock(t) * Treatment	3.42	2.89	0.23	0.33	0.27	0.37
	(3.06)	(3.04)	(1.10)	(1.10)	(1.12)	(1.11)
Shock(t-1) * Treatment		4.55*		2.11		2.02**
		(2.41)		(1.36)		(0.88)
Treatment	-0.75	-1.47	-0.13	-0.59	-0.25	-1.59
	(0.82)	(0.98)	(0.64)	(0.69)	(0.94)	(1.14)
N	2477	2477	2477	2477	2477	2477
Average size	7.03	7.03	7.03	7.03	7.03	7.03

OLS estimates. *, **, *** indicate significance at the 10, 5 and 1%. SE clustered by village. Column 1-2 represent the effect of rainfall shocks defined as absolute deviation from long-run average. Column 3-4 represent the effect of rainfall shocks defined as a dummy if absolute deviation is 1 standard deviation above or below long-run average. Column 5-6 represent the effect of absolute deviation of rainfall from long-run average divided by standard deviation. Covariates: HH head age, age-squared, gender, work, education, HH poor, size, composition, and state FE. "Treatment" denotes cash transfer from PROGRESA.

Table 4.8 shows the effect of rainfall shocks on the clustering coefficient calculated by household and network level. The effect of rainfall shocks is negative in the control group and positive for lagged interaction terms in the treatment group. However, the effects are very small and not significant.

4.5.2 Mechanisms

We find a smaller degree and size of the extended family network within control villages where rainfall shocks occurred. These negative effects are counteracted by cash transfers provided by PROGRESA, as shown in positive interaction effects on treatment villages with rainfall shocks. To explore what derived these results, we study the relationships between rainfall shocks and a household's decision to split for marriage or to migrate.

Table 4.7: The effect of rainfall shock on average path length

	Average path length					
	(1)	(2)	(3)	(4)	(5)	(6)
Rainfall shock	Abs deviation		1sd above/below avg		Abs dev/sd	
Shock (t)	0.01	0.01	0.03	0.03	0.04	0.04
	(0.06)	(0.06)	(0.04)	(0.04)	(0.04)	(0.03)
Shock (t-1)		-0.06		0.01		-0.00
		(0.08)		(0.04)		(0.03)
Shock(t) * Treatment	0.01	0.02	-0.02	-0.02	-0.01	-0.01
	(0.10)	(0.10)	(0.05)	(0.04)	(0.04)	(0.04)
Shock (t-1) * Treatment		-0.01		-0.07		-0.02
		(0.10)		(0.05)		(0.04)
Treatment	-0.03	-0.03	-0.02	-0.01	-0.02	-0.01
	(0.03)	(0.03)	(0.02)	(0.03)	(0.03)	(0.04)
N	2477	2477	2477	2477	2477	2477
Mean avg. path length	1.44	1.44	1.44	1.44	1.44	1.44

OLS estimates. *, **, *** indicate significance at the 10, 5 and 1%. SE clustered by village. Column 1-2 represent the effect of rainfall shocks defined as absolute deviation from long-run average. Column 3-4 represent the effect of rainfall shocks defined as a dummy if absolute deviation is 1 standard deviation above or below long-run average. Column 5-6 represent the effect of absolute deviation of rainfall from long-run average divided by standard deviation. Covariates: HH head age, age-squared, gender, work, education, HH poor, size, composition, and state FE. "Treatment" denotes cash transfer from PROGRESA.

Split households can affect within-village extended family network degree or size by creating new nodes in the network. Some household members may choose to leave to form a new household in the village, which is typically the case for young couples who initially stayed with their in-laws after marriage while they saved to move out. Therefore, leaving the household for marriage within the village can increase the extended family network degree and size. Female household members could move to marry another household within the village. Moreover, female household members might move outside the village for marriage. However, these decisions of female household members will not change within-village networks. Furthermore, households may migrate out of the village for work. If these migrations move the whole household out of the village, the extended family network degree and size could be reduced by removing a node from the networks.

Thus, it is worthwhile to examine if rainfall shocks affected household members leaving

Table 4.8: The effect of rainfall shock on clustering coefficient

	Household level clustering coefficient					
	(1)	(2)	(3)	(4)	(5)	(6)
Rainfall shock	Abs deviation		1sd above/below avg		Abs dev/sd	
Shock (t)	-0.01	-0.01	-0.01	-0.01	-0.01	-0.02
	(0.03)	(0.03)	(0.01)	(0.01)	(0.01)	(0.01)
Shock (t-1)		0.02		-0.01		-0.01
		(0.03)		(0.02)		(0.01)
Shock(t) * Treatment	-0.02	-0.03	0.00	0.00	-0.00	0.00
	(0.03)	(0.03)	(0.02)	(0.02)	(0.01)	(0.01)
Shock (t-1) * Treatment		0.02		0.03		0.02
		(0.04)		(0.02)		(0.02)
Treatment	0.01	0.00	0.00	-0.01	0.00	-0.02
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)
N	12339	12339	12339	12339	12339	12339
Avg. clustering coeff.	0.71	0.71	0.71	0.71	0.71	0.71
	Network level clustering coefficient					
	(1)	(2)	(3)	(4)	(5)	(6)
Rainfall shock	Abs deviation		1sd above/below avg		Abs dev/sd	
Shock (t)	-0.06	-0.06	-0.02	-0.02	-0.01	-0.01
	(0.07)	(0.07)	(0.03)	(0.03)	(0.03)	(0.03)
Shock (t-1)		-0.03		-0.02		-0.01
		(0.06)		(0.04)		(0.02)
Shock(t) * Treatment	-0.00	-0.02	0.01	0.02	-0.01	-0.00
	(0.09)	(0.09)	(0.04)	(0.04)	(0.04)	(0.04)
Shock(t-1) * Treatment		0.11		0.06		0.03
		(0.09)		(0.05)		(0.03)
Treatment	0.02	-0.00	0.01	-0.00	0.02	-0.00
	(0.03)	(0.03)	(0.02)	(0.02)	(0.03)	(0.04)
N	1413	1413	1413	1413	1413	1413
Avg. clustering coeff.	0.58	0.58	0.58	0.58	0.58	0.58

OLS estimates. *, **, *** indicate significance at the 10, 5 and 1%. SE clustered by village. Column 1-2 represent the effect of rainfall shocks defined as absolute deviation from long-run average. Column 3-4 represent the effect of rainfall shocks defined as a dummy if absolute deviation is 1 standard deviation above or below long-run average. Column 5-6 represent the effect of absolute deviation of rainfall from long-run average divided by standard deviation. Covariates: HH head age, age-squared, gender, work, education, HH poor, size, composition, and state FE. "Treatment" denotes cash transfer from PROGRESA.

the household for marriage or whether households migrate out of village. Cash transfers might relieve liquidity constraints from rainfall shocks and change the effects. The effects of these different actions can be estimated through the model 4.1 changing dependent variables to split household or migration used with the same control variables.

Split households for marriage

Table 4.9: The effect of rainfall shock on leaving household for marriage

Anyone aged 15-39 left household for marriage within village in 12 months						
	(1)	(2)	(3)	(4)	(5)	(6)
Rainfall shock / 100	Abs deviation		1sd above/below avg		Abs deviation/sd	
Shock(t)	-0.71	-0.74*	-0.39	-0.42	-0.18	-0.24
	(0.45)	(0.44)	(0.29)	(0.29)	(0.24)	(0.25)
Shock(t-1)		-0.75		-0.29		-0.37*
		(0.58)		(0.32)		(0.22)
Shock(t) * Treatment	0.23	0.41	-0.12	-0.15	-0.06	-0.04
	(0.58)	(0.56)	(0.37)	(0.37)	(0.31)	(0.31)
Shock(t-1) * Treatment		-0.45		-0.33		-0.26
		(0.83)		(0.40)		(0.32)
Treatment	0.00	0.00	0.00	0.00	0.00	0.00
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)

OLS estimates. *, **, *** indicate significance at the 10, 5 and 1%. SE clustered by village. Column 1-2 represent the effect of rainfall shocks defined as absolute deviation from long-run average. Column 3-4 represent the effect of rainfall shocks defined as a dummy if absolute deviation is 1 standard deviation above or below long-run average. Column 5-6 represent the effect of absolute deviation of rainfall from long-run average divided by standard deviation. Covariates: HH head age, age-squared, gender, work, education, HH poor, size, composition, and state FE. "Treatment" denotes cash transfer from PROGRESA.

Table 4.9 presents the effect of rainfall shock on split households for marriage. The indicator of split households is a dummy variable of whether anyone aged 15–39 left one's household within the village 12 months prior to the survey. This variable does not include those who migrated out of the village. The rainfall shocks reduce the likelihood of a household member leaving the household for marriage in the control group. However, the effects are noisy and inconsistent by specifications and only statistically significant in the case of especially large shocks. The result finds that a one-metre increase in the absolute deviation of rainfall from the long-run average decreases the probability of anyone leaving the household for marriage by 0.74 percentage points. Unlike the effects on network degree and size, cash transfers do not counter the negative effect of rainfall shocks.

Table 4.10 represents the effect of rainfall shocks on leaving households for marriage, esti-

Table 4.10: The effect of rainfall shock on leaving HH for marriage by gender

Any female aged 15-39 left household for marriage within village in 12 months						
	(1)	(2)	(3)	(4)	(5)	(6)
Rainfall shock / 100	Abs deviation		1sd above/below avg		Abs deviation/sd	
Shock(t)	-0.75**	-0.77**	-0.38*	-0.40*	-0.25	-0.29
	(0.36)	(0.35)	(0.23)	(0.23)	(0.19)	(0.20)
Shock(t-1)		-0.44		-0.13		-0.26
		(0.46)		(0.27)		(0.19)
Shock(t) * Treatment	0.53	0.66	0.15	0.13	0.16	0.18
	(0.49)	(0.48)	(0.30)	(0.30)	(0.25)	(0.25)
Shock(t-1) * Treatment		-0.45		-0.19		-0.12
		(0.62)		(0.34)		(0.27)
Treatment	-0.00	0.00	0.00	0.00	-0.00	0.00
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Any male aged 15-39 left household for marriage within village in 12 months						
	(1)	(2)	(3)	(4)	(5)	(6)
Rainfall shock / 100	Abs deviation		1sd above/below avg		Abs deviation/sd	
Shock(t)	0.10	0.09	-0.07	-0.08	0.04	0.01
	(0.30)	(0.29)	(0.17)	(0.17)	(0.14)	(0.14)
Shock(t-1)		-0.28		-0.17		-0.14
		(0.29)		(0.17)		(0.12)
Shock(t) * Treatment	-0.21	-0.15	-0.18	-0.21	-0.17	-0.16
	(0.38)	(0.38)	(0.23)	(0.23)	(0.18)	(0.18)
Shock(t-1) * Treatment		-0.12		-0.25		-0.19
		(0.52)		(0.22)		(0.16)
Treatment	0.00*	0.00	0.00	0.00*	0.00*	0.00*
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)

OLS estimates. *, **, *** indicate significance at the 10, 5 and 1%. SE clustered by village. Column 1-2 represent the effect of rainfall shocks defined as absolute deviation from long-run average. Column 3-4 represent the effect of rainfall shocks defined as a dummy if absolute deviation is 1 standard deviation above or below long-run average. Column 5-6 represent the effect of absolute deviation of rainfall from long-run average divided by standard deviation. Covariates: HH head age, age-squared, gender, work, education, HH poor, size, composition, and state FE. "Treatment" denotes cash transfer from PROGRESA.

mated for females and males separately. Rainfall shocks significantly reduced the probability of any females who left the household for marriage within 12 months in the control group. Lagged rainfall shocks do not have significant effects, and cash transfer does not counter the negative effect of rainfall shocks. The size and significance of the effects are larger for females than those for anyone (females and males) leaving households for marriage. The result finds that a one-metre increase in the absolute deviation of rainfall from the long-run average decreases the probability of any female leaving the household for marriage by 0.77 percentage points. Moreover, the absolute deviation of rainfall 1 standard deviation above or below long-term average decreases the probability of any female leaving the household for marriage by 0.4 percentage points. The effects are not significant for any male leaving the household for marriage.

It is common for females to leave their households after marriage and move to other households because of the culture from patriotic society in rural Mexico. Indeed, 68% of households with any female who left her household moved to other household for marriage in the main sample. This local marriage norms are reflected in the gender differences in the effect of rainfall shock on split households for marriage.

In summary, fewer females leaving their household for marriage account for fewer split households after rainfall shocks in control villages. These findings suggest that fewer split households due to rainfall shocks may cause smaller extended family network degree and size. Furthermore, no significant interaction effects on split households in treatment villages reflect that cash transfers do not counter the effect of rainfall shocks on split households.

Migration

Migration has been argued to be the primary method of consumption smoothing in developing countries. Given the variable income in rural areas, households have incentives to pool risk through temporary migration or migration for marriage (Rosenzweig and Stark, 1989; Morten, 2016). This mechanism may lead to more migration from villages with rainfall shocks. Therefore, migration is another possible factor that caused the decrease in within-village extended family network degree and size because a household will be removed from the extended family network within the village if all household members migrate out of the village. Even if not all members of households migrate, the remaining household members of migrants might dissolve households or combine with other households because family members of migrants want to relieve credit constraints by resource pooling for improved access to remittances from the migrants or to supplement family employment who could replace the migrants (Bertoli and Murard, 2020).

Table 4.11: The effect of rainfall shock on work migration

Anyone aged 15-39 migrated for work out of village in 12 months						
	(1)	(2)	(3)	(4)	(5)	(6)
Rainfall shock / 100	Abs deviation		1sd above/below avg		Abs deviation/sd	
Shock (t)	0.80	0.78	0.26	0.16	0.17	0.07
	(1.12)	(1.12)	(0.82)	(0.82)	(0.55)	(0.55)
Shock (t-1)		-0.43		-1.15*		-0.62
		(1.26)		(0.63)		(0.52)
Shock (t) * Treatment	-2.57**	-2.47**	-0.78	-0.83	-0.78	-0.75
	(1.22)	(1.24)	(1.05)	(1.06)	(0.72)	(0.72)
Shock (t-1) * Treatment		-0.26		-0.59		-0.72
		(1.84)		(0.90)		(0.68)
Treatment	0.01	0.01	0.00	0.01	0.01	0.01
	(0.00)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
N	16399	16399	16399	16399	16399	16399

OLS estimates. *, **, *** indicate significance at the 10, 5 and 1%. SE clustered by village. Column 1-2 represent the effect of rainfall shocks defined as absolute deviation from long-run average. Column 3-4 represent the effect of rainfall shocks defined as a dummy if absolute deviation is 1 standard deviation above or below long-run average. Column 5-6 represent the effect of absolute deviation of rainfall from long-run average divided by standard deviation. Covariates: HH head age, age-squared, gender, work, education, HH poor, size, composition, and state FE. "Treatment" denotes cash transfer from PROGRESA.

Table 4.12: The effect of rainfall shock on marriage migration

Anyone aged 15-39 migrated for work out of village in 12 months						
	(1)	(2)	(3)	(4)	(5)	(6)
Rainfall shock / 100	Abs deviation		1sd above/below avg		Abs deviation/sd	
Shock (t)	-0.04	-0.05	-0.06	-0.08	0.03	0.00
	(0.59)	(0.58)	(0.30)	(0.29)	(0.24)	(0.23)
Shock (t-1)		-0.53		-0.28		-0.17
		(0.66)		(0.31)		(0.28)
Shock (t) * Treatment	-0.27	-0.34	-0.06	-0.05	-0.07	-0.03
	(0.62)	(0.62)	(0.39)	(0.39)	(0.29)	(0.28)
Shock (t-1) * Treatment		1.21		0.08		0.33
		(1.04)		(0.41)		(0.40)
Treatment	0.00	-0.00	0.00	0.00	0.00	-0.00
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
N	16399	16399	16399	16399	16399	16399

OLS estimates. *, **, *** indicate significance at the 10, 5 and 1%. SE clustered by village. Column 1-2 represent the effect of rainfall shocks defined as absolute deviation from long-run average. Column 3-4 represent the effect of rainfall shocks defined as a dummy if absolute deviation is 1 standard deviation above or below long-run average. Column 5-6 represent the effect of absolute deviation of rainfall from long-run average divided by standard deviation. Covariates: HH head age, age-squared, gender, work, education, HH poor, size, composition, and state FE. "Treatment" denotes cash transfer from PROGRESA.

Two common reasons for migrations are work and marriage. Table 4.11 presents the effect of rainfall shocks on work migration. The effect of rainfall shocks is positive in the control group, but the lagged effects are negative. However, the effects are not significant except for the large effect. Although interaction terms are negative for all specifications, only the absolute deviation of rainfall from the long-run average decreases work migration in the treatment group. This finding suggests that those who needed less money after receiving cash transfers may have less incentive to migrate and work out of the village. In addition, the conditions attached to the cash transfer could have prevented the beneficiaries from leaving their villages. The effects of rainfall shock on marriage migration are not significant for both the control and treatment groups (Table 4.12). Less work migration influenced by rainfall shocks infers how cash transfers counteracted negative effects of rainfall shocks on extended family network degree and size in treatment villages. This finding is due to the lower likelihood of entire households migrating out of the village or migrant households dissolving and combining with other households after receiving cash transfers.

4.6 Robustness

We re-estimate the effects of rainfall shocks with different definitions of the phenomena to assess the robustness of the results. Two different definitions of rainfall shocks are generated. Firstly, rainfall shock variables in year t-2 are included as additional variables. In model 4.1, lagged rainfall shocks with a lag of one year were included to consider possible time differences of the effect. As a robustness check, rainfall shocks in both years t-1 and t-2 are included.

The effects of rainfall shocks on the degree are negative in control villages and interaction effects are positive in treatment villages (Table 4.13). Likewise, the effects of rainfall shocks on network size are negative in the control group and interaction effects are positive in the treatment group (Table 4.14). The effects of the clustering coefficient and average path length are not significant (Table 4.19 and 4.20). The effects of lagged rainfall shocks with a lag of two years are not significant for all types of network characteristics. The effect of

Table 4.13: The effect of rainfall shock on degree

Rainfall shock	Abs deviation	1sd above/below avg	Abs deviation/sd
Shock(t)	-1.90**	-0.85	-0.85
	(0.86)	(0.76)	(0.62)
Shock(t-1)	-0.59	-0.84	-0.59
	(1.14)	(0.55)	(0.41)
Shock(t-2)	-0.98	-0.11	-0.82
	(1.36)	(0.65)	(0.67)
Shock(t) * Treatment	0.83	0.08	0.16
	(0.96)	(0.80)	(0.67)
Shock(t-1) * Treatment	3.41**	1.20*	1.12**
	(1.35)	(0.67)	(0.47)
Shock(t-2) * Treatment	-0.34	0.01	0.65
	(1.45)	(0.74)	(0.72)
Treatment	-0.97	-0.52	-1.55
	(0.65)	(0.71)	(1.14)
N	16399	16399	16399
Average degree	4.11	4.11	4.11

OLS estimates. *, **, *** indicate significance at the 10, 5 and 1%. SE clustered by village. Covariates: HH head age, age-squared, gender, work, education, HH poor, size, composition, and state FE.

Table 4.14: The effect of rainfall shock on size

Rainfall shock	Abs deviation	1sd above/below avg	Abs deviation/sd
Shock(t)	-2.88**	-0.59	-0.69
	(1.36)	(0.84)	(0.75)
Shock(t-1)	-1.89	-1.15	-1.00
	(1.67)	(0.92)	(0.74)
Shock(t-2)	0.29	1.10	0.25
	(1.78)	(1.00)	(0.94)
Shock(t) * Treatment	2.78	0.27	0.37
	(3.04)	(1.09)	(1.10)
Shock(t-1) * Treatment	5.55*	2.38*	2.11**
	(3.09)	(1.41)	(1.01)
Shock(t-2) * Treatment	-1.91	-0.72	-0.19
	(2.39)	(1.24)	(1.16)
Treatment	-1.25	-0.41	-1.50
	(1.03)	(0.72)	(1.24)
N	2477	2477	2477
Average size	7.03	7.03	7.03

OLS estimates. *, **, *** indicate significance at the 10, 5 and 1%. SE clustered by village. Covariates: HH head age, age-squared, gender, work, education, HH poor, size, composition, and state FE.

rainfall shocks on network characteristics is robust with lagged rainfall shocks with a lag of two years.

Next, rainfall shocks are defined by different benchmarks, using long-run average rainfall

Table 4.15: Rainfall statistics

	N	Mean	SD
Rainfall at year t	16399	1243.70	742.40
Rainfall at year t-1	16399	1051.74	651.25
10-year average	16399	1125.02	648.74
10-year sd	16399	291.18	183.49
Absolute deviation from 10y avg / 1000 at t	16399	0.22	0.27
Absolute deviation from 10y avg / 1000 at t-1	16399	0.19	0.21
Absolute deviation 1sd above/below 10y avg at t	16399	0.33	0.47
Absolute deviation 1sd above/below 10y avg at t-1	16399	0.29	0.45
Absolute deviation from 10y avg / sd at t	16399	0.87	0.96
Absolute deviation from 10y avg / sd at t-1	16399	0.67	0.51

as ten-year average instead of a 15-year average. Table 4.15 presents rainfall statistics generated with ten-year average rainfall. The ten-year average rainfall is approximately 1,125mm, and the ten-year standard deviation is 291mm, smaller than the 15-year standard deviation (305mm). Rainfall shocks are defined as (a) absolute deviation from the ten-year average rainfall, (b) whether the absolute deviation is one standard deviation above or below the ten-year average, and (c) absolute deviation from the ten-year average divided by the ten-year standard deviation. The measures are similar to those generated with the 15-year average rainfall. The absolute deviation divided by the standard deviation (0.87) is larger than that created with the 15-year average rainfall (0.68) due to the smaller standard deviation.

Table 4.16: The effect of rainfall shock on degree

Rainfall shock from 10y avg	Abs deviation		1sd above/below avg		Abs deviation/sd	
Shock(t)	-2.41**	-2.22**	-0.91	-0.95	-0.57	-0.55
	(1.20)	(1.10)	(0.79)	(0.80)	(0.46)	(0.45)
Shock(t-1)		-1.29		-0.34		-0.57
		(1.12)		(0.64)		(0.65)
Shock(t) * Treatment	1.94	1.56	-0.02	0.03	0.43	0.36
	(1.23)	(1.16)	(0.77)	(0.78)	(0.48)	(0.47)
Shock(t-1) * Treatment		2.70**		1.08		1.03
		(1.29)		(0.74)		(0.71)
Treatment	-0.63	-1.07	-0.24	-0.55	-0.60	-1.22
	(0.56)	(0.71)	(0.51)	(0.68)	(0.63)	(1.00)
N	16399	16399	16399	16399	16399	16399
Average degree	4.11	4.11	4.11	4.11	4.11	4.11

OLS estimates. *, **, *** indicate significance at the 10, 5 and 1%. SE clustered by village.
Covariates: HH head age, age-squared, gender, work, education, HH poor, composition, and state FE.

The absolute deviation of rainfall from the ten-year average rainfall negatively affects the network degree in the control group but has a positive interaction effect on the treatment group (Table 4.16). The absolute deviation of the rainfall from the ten-year average also reduces network size in the control group but does not have a significant interaction effect in the treatment group (Table 4.17). The effect of rainfall shock variables generated with the ten-year average does not have a significant influence on the clustering coefficient and average path length (Table 4.21 and 4.22). Therefore, the negative effects of rainfall shocks on network degree and size in the control group are robust with rainfall shocks compared to the ten-year average rainfall.

Table 4.17: The effect of rainfall shock on size

Rainfall shock from 10y avg	Abs deviation		1sd above/below avg		Abs deviation/sd	
Shock(t)	-2.66**	-2.51**	-1.05	-1.01	-0.43	-0.41
	(1.19)	(1.18)	(0.90)	(0.88)	(0.43)	(0.43)
Shock(t-1)		-0.83		0.41		-0.23
		(1.30)		(0.84)		(0.64)
Shock(t) * Treatment	3.28	2.90	0.11	0.06	0.69	0.59
	(2.90)	(3.00)	(1.07)	(1.05)	(0.81)	(0.82)
Shock(t-1) * Treatment		2.00		0.93		0.98
		(2.23)		(1.19)		(0.97)
Treatment	-0.81	-1.10	-0.11	-0.29	-0.66	-1.20
	(0.86)	(0.94)	(0.69)	(0.77)	(0.87)	(1.02)
N	2477	2477	2477	2477	2477	2477
Average size	7.03	7.03	7.03	7.03	7.03	7.03

OLS estimates. *, **, *** indicate significance at the 10, 5 and 1%. SE clustered by village.
Covariates: HH head age, age-squared, gender, work, education, HH poor, composition and state FE.

Finally, the effects of rainfall riskiness on within-village extended family network statistics are estimated to test for strategic network formation in response to risk. The measure of rainfall riskiness is the coefficient of variation, the 15-year standard deviation divided by the 15-year average. Rainfall riskiness did not have a significant impact on any network characteristics (Table 4.18). Thus, households did not change their network structure due to the long-term weather risks they face in their villages.

Table 4.18: The effect of rainfall riskiness

	Degree	Clustering coefficient	Size	Average path length
Coefficient of variation	-0.02	-0.01	2.56	-0.07
	(1.47)	(0.03)	(0.08)	(2.48)
N	16399	12339	2477	2477

OLS estimates. *, **, *** indicate significance at the 10, 5 and 1%. SE clustered by village. Covariates: HH head age, age-squared, gender, work, education, HH poor, size, composition, and state FE.

4.7 Conclusion

Through interactions in extended family networks within villages, risk and resource sharing are important for lives in developing countries. We studied the effect of rainfall shock on within-village extended family networks to analyse the reactions of households to the variation of agricultural incomes caused by weather shocks. We used rich data from a household panel survey collected to assess a poverty alleviation programme named PROGRESA to estimate the effect of rainfall shocks on extended family networks in rural Mexico.

Exploiting data on the paternal and maternal surnames of household heads and spouses and the naming traditions of Mexico, family links of each household to others in the village could be identified. Rainfall data are matched at the village level, and rainfall shocks are defined as the absolute deviation of rainfall from the long-run average and whether the rainfall amount is one standard deviation above or below the long-run village mean. Using the exogeneity of rainfall shock and treatment status for cash transfer in PROGRESA, the effects of rainfall shocks on within-village extended family network statistics allowing for treatment effects are estimated. Extended family networks have smaller network degrees and sizes in control villages where rainfall shocks have occurred. However, the negative effects of rainfall shocks are countervailed in the treatment group with cash transfers. Further analysis finds that females are less likely to leave households within villages for marriage in the control group, suggesting a decrease of split households in control villages with rainfall shocks, related to a smaller degree and size in extended family networks. Cash transfer countervailing the reduction of network degree and size in treatment group might have resulted from the negative interaction effect of rainfall shocks on work migration.

Rainfall shocks not only reduced agricultural income in rural areas of developing countries but also significantly changed within-village extended family network structures. This change is likely to have resulted from decisions not to split households. For example, those who experienced rainfall shocks and married were less likely to leave their households and create new households. Households may have preferred to aggregate income and take advantage of economies of scale to address income shocks. Insignificant effects of rainfall shocks on marriage migration imply that households in rural Mexico did not necessarily deal with the income variation from rainfall shocks by consumption smoothing expected from migration for the marriage. In addition, cash transfer played a crucial role in changing the effect of rainfall shocks. This finding emphasises that financial transfers can mitigate liquidity constraints caused by income shocks. Cash transfers reduced work migration but did not prevent split households. This study shows the importance of within-village extended family network structures that reflect developing countries' risk-coping mechanisms. Furthermore, the findings point out the possible roles of financial transfers in alleviating income shocks in rural regions.

Appendix 4A: Definitions of network statistics

- Degree: The number of connected nodes (households in this analysis) to each node in a network (within-village extended family network in this case), showing how well-connected a node is
- Betweenness: The number of shortest paths that go through a given node
- Closeness: The number of steps needed to access every other node from a given node
- Eigenvector centrality: Values of the first eigenvector of the adjacency matrix for network
- Clustering coefficient: A probability that adjacent nodes of each node are connected, which is an indicator of transitivity
- Size: The number of nodes in each network
- Density: The number of edges that exist divided by the number of edges that are possible
- Average path length: The average number of steps along the shortest paths for all existing pairs of nodes
- Diameter: the length of the longest path (in number of edges) between two nodes

Appendix 4B: Tables

Table 4.19: The effect of rainfall shock on clustering coefficient

Rainfall shock	Abs deviation	1sd above/below avg	Abs deviation/sd
Shock(t)	-0.01	-0.01	-0.02
	(0.03)	(0.01)	(0.01)
Shock(t-1)	0.03	-0.01	-0.01
	(0.04)	(0.02)	(0.01)
Shock(t-2)	-0.03	-0.01	-0.01
	(0.03)	(0.01)	(0.01)
Shock(t) * Treatment	-0.04	0.00	-0.00
	(0.03)	(0.02)	(0.01)
Shock(t-1) * Treatment	0.07	0.03	0.03
	(0.05)	(0.02)	(0.02)
Shock(t-2) * Treatment	-0.08*	-0.01	-0.02
	(0.05)	(0.02)	(0.02)
Treatment	0.01	-0.01	-0.01
	(0.01)	(0.01)	(0.02)
N	12339	12339	12339

OLS estimates. *, **, *** indicate significance at the 10, 5 and 1%. SE clustered by village.
Covariates: HH head age, age-squared, gender, work, education, HH size, composition, state FE

Table 4.20: The effect of rainfall shock on average path length

Rainfall shock	Abs deviation	1sd above/below avg	Abs deviation/sd
Shock(t)	0.02	0.04	0.04
	(0.06)	(0.04)	(0.03)
Shock(t-1)	-0.14	0.00	-0.03
	(0.09)	(0.04)	(0.03)
Shock(t-2)	0.12	0.03	0.06
	(0.09)	(0.04)	(0.04)
Shock(t) * Treatment	0.03	-0.02	-0.00
	(0.10)	(0.04)	(0.04)
Shock(t-1) * Treatment	-0.01	-0.06	-0.01
	(0.11)	(0.05)	(0.04)
Shock(t-2) * Treatment	0.03	0.06	0.01
	(0.11)	(0.05)	(0.05)
Treatment	-0.04	-0.02	-0.02
	(0.04)	(0.03)	(0.05)
N	2477	2477	2477

OLS estimates. *, **, *** indicate significance at the 10, 5 and 1%. SE clustered by village.
Covariates: HH head age, age-squared, gender, work, education, HH size, composition, state FE

Table 4.21: The effect of rainfall shock on clustering coefficient

Rainfall shock from 10y avg	Abs deviation		1sd above/below avg		Abs deviation/sd	
Shock(t)	-0.01	-0.01	-0.01	-0.01	-0.01	-0.01
	(0.03)	(0.03)	(0.01)	(0.01)	(0.01)	(0.01)
Shock(t-1)		0.01		-0.00		-0.01
		(0.03)		(0.01)		(0.01)
Shock(t) * Treatment	-0.02	-0.02	-0.00	-0.00	0.00	-0.00
	(0.03)	(0.03)	(0.02)	(0.02)	(0.01)	(0.01)
Shock(t-1) * Treatment		0.02		0.03		0.02
		(0.04)		(0.02)		(0.02)
Treatment	0.00	0.00	0.00	-0.01	0.00	-0.01
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)
N	12339	12339	12339	12339	12339	12339

OLS estimates. *, **, *** indicate significance at the 10, 5 and 1%. SE clustered by village.
Covariates: HH head age, age-squared, gender, work, education, HH poor, composition, and state FE.

Table 4.22: The effect of rainfall shock on average path length

Rainfall shock from 10y avg	Abs deviation		1sd above/below avg		Abs deviation/sd	
Shock(t)	0.01	0.02	0.03	0.03	0.02	0.02
	(0.06)	(0.06)	(0.04)	(0.04)	(0.02)	(0.02)
Shock(t-1)		-0.06		-0.01		-0.02
		(0.07)		(0.04)		(0.03)
Shock(t) * Treatment	-0.00	-0.01	0.01	0.01	-0.02	-0.02
	(0.10)	(0.10)	(0.04)	(0.04)	(0.03)	(0.03)
Shock(t-1) * Treatment		0.04		-0.04		0.00
		(0.09)		(0.04)		(0.04)
Treatment	-0.03	-0.04	-0.03	-0.02	-0.01	-0.01
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.04)
N	2477	2477	2477	2477	2477	2477

OLS estimates. *, **, *** indicate significance at the 10, 5 and 1%. SE clustered by village.
Covariates: HH head age, age-squared, gender, work, education, HH poor, composition, and state FE.

Chapter 5

Conclusion

In this thesis, social networks in developing countries are studied using peer effects on behaviour change in religious groups to obtain HIV test results, the influence of village neighbours' subjective expectations about HIV and the impact of rainfall shocks on the extended family network structure. The first chapter examined peer effects in religious groups on learning about HIV test results. An important goal for HIV prevention is to overcome social stigma and cause more people to get tested and learn about their HIV status. Religion and religious networks can be useful tools to approach HIV prevention in sub-Saharan Africa. I applied the data from rural Malawi and used randomly assigned monetary incentives as instrumental variables for the portion of religious groups who learned about HIV test results. The IV estimates generated positive and significant peer effects for religious group members' obtaining test results on individuals obtaining their personal test results. This effect is stronger than the effect examined from geographical neighbours. This finding suggests the possibility of using the role of religious networks for HIV prevention in sub-Saharan Africa. Policies may consider the possibility of assisting religious networks in overcoming social stigma and scaling up health interventions in developing countries. HIV prevention can target religious groups to efficiently increase HIV testing via peer effects with limited resources.

The second chapter studies peer effects on the subjective expectations regarding an individual's HIV infection. Peer effects might be useful to deal with HIV prevention in developing countries where information provision and social learning can play a crucial role. The data includes extensive subjective expectation data collected by visual and easy methods from the respondents. I estimate the peer effects of village neighbours on individuals' subjective likelihood of HIV infection influenced by HIV testing. Randomised monetary incentives to encourage learning about test results are used as the source of variation to identify peer effects. I find significant and positive peer effects regarding village neighbours on an individual's subjective likelihood of current HIV infection and HIV infection in the future. Marginal treatment effects show the heterogeneity of peer effects depending on the subjective expectations of neighbours. Significant effect of neighbours with a low subjective likelihood of HIV infection implies that those who learned about test results have larger peer effects than others. Therefore, peer effects can effectively correct subjective expectations and provide relevant information for HIV risk and prevention in sub-Saharan Africa. Larger peer effects for those who learned about their HIV test results emphasise the need to promote HIV testing.

Through interactions in extended family networks within villages, risk and resource sharing are important for lives in developing countries. The third chapter examined the effect of rainfall shock on within-village extended family networks to analyse the reactions of households to the variation of agricultural incomes caused by weather shocks. Using rich data from a household panel survey collected for the poverty mitigation programme titled PROGRESA in rural Mexico, family links of each household to others in the village could be identified. Extended family networks have smaller network degrees and sizes in control villages where rainfall shocks have occurred. Cash transfers counteracted the negative effect of rainfall shocks on network degree and size. The possible drivers of these features are fewer household members leaving households within control villages and less work migration in treatment villages. The results suggest households may prefer to aggregate income and take advantage of economies of scale to address income shocks. However, financial transfers can mitigate liquidity constraints caused by income shocks by reducing work migration. This finding shows the importance of within-village extended family network structures that reflect developing countries' risk-coping mechanisms.

The findings in this thesis emphasise the importance of social networks in developing countries in terms of behaviour, belief and risk-sharing. Policies exploiting these characteristics of networks will efficiently yield benefits, such as diffusing beneficial health behaviour, affecting subjective beliefs and relieving liquidity constraints caused by adverse income shocks. Further studies could be conducted on the channels through which peer effects overcome social stigma, subjective beliefs change the behaviour, or resources are shared among networks. More analysis of the mechanisms using detailed network data will provide additional interesting aspects to studies on networks and suggest policy implications in developing countries.

Bibliography

[1] Adams, J., and Trinitapoli, J. The Malawi religion project: Data collection and selected analyses. *Demographic Research 21*, 10 (2009), 255–288.

[2] Adhvaryu, A., Nyshadham, A., Molina, T., and Tamayo, J. Helping children catch up: Early life shocks and the progresa experiment. Working Paper 24848, National Bureau of Economic Research, 2018.

[3] Algan, Y., Dalvit, N., Do, Q.-A., Chapelain, A. L., and Zenou, Y. Friendship networks and political opinions: A natural experiment among future French politicians. Sciences Po publications 13771, Sciences Po, 2019.

[4] Andresen, M. E. Exploring marginal treatment effects: Flexible estimation using stata. *Stata Journal 18*, 1 (2018), 118–158.

[5] Angelucci, M. Aid and migration: An analysis of the impact of progresa on the timing and size of labour migration. *IZA Institute of Labor Economics Discussion Paper Series* (2004).

[6] Angelucci, M., De Giorgi, G., Rangel, M. A., and Rasul, I. Extended family networks in rural mexico: A descriptive analysis. IZA Discussion Papers 4498, Institute of Labor Economics (IZA), 2009.

[7] Angelucci, M., De Giorgi, G., Rangel, M. A., and Rasul, I. Family networks and school enrolment: Evidence from a randomized social experiment. *Journal of Public Economics 94*, 3 (2010), 197–221.

[8] Angelucci, M., De Giorgi, G., and Rasul, I. Consumption and investment in resource pooling family networks. *The Economic Journal 128*, 615 (2017), 2613–2651.

[9] Angelucci, M., Giorgi, G. D., Rangel, M., and Rasul, I. Village economies and the structure of extended family networks. *The B.E. Journal of Economic Analysis & Policy 9*, 1 (2009).

[10] Anglewicz, P. Migration, marital change, and hiv infection in malawi. *Demography 49*, 1 (2012), 239–265.

[11] Anglewicz, P., and Kohler, H.-P. Overestimating HIV infection:: The construction and accuracy of subjective probabilities of HIV infection in rural Malawi. *Demographic Research 20*, 6 (2009), 65–96.

[12] Angrist, J. D. The perils of peer effects. *Labour Economics 30* (2014), 98–108.

[13] Attanasio, O. P. Expectations and perceptions in developing countries: Their measurement and their use. *American Economic Review 99*, 2 (2009), 87–92.

[14] Badham, J., Kee, F., and Hunter, R. F. Simulating network intervention strategies: Implications for adoption of behaviour. *Network Science 6*, 2 (2018), 265–280.

[15] Banerjee, A., Breza, E., G a, A., Duflo, E., Jackson, M., and Kinnan, C. Changes in social network structure in response to exposure to formal credit markets. *SSRN Electronic Journal* (2021).

[16] Banerjee, A., Chandrasekhar, A. G., Duflo, E., and Jackson, M. O. The diffusion of microfinance. Working Paper 17743, National Bureau of Economic Research, 2012.

[17] Bank, W. *Malawi at a Glance*. World Bank, Washington, DC, 2013.

[18] Becerra-Valbuena, L. G., and Millock, K. Gendered migration responses to drought in malawi. *Journal of Demographic Economics 87*, 3 (2021), 437–477.

[19] Becker, M. H., Drachman, R. H., and Kirscht, J. P. A new approach to explaining sick-role behavior in low-income populations. *American Journal of Public Health 64*, 3 (1974), 205–216.

[20] Benagiano, G., Carrara, S., Filippi, V., and Brosens, I. Condoms, hiv and the roman catholic church. *Reproductive BioMedicine Online 22*, 7 (2011), 701–709.

[21] Bertoli, S., and Murard, E. Migration and co-residence choices: Evidence from Mexico. *Journal of Development Economics 142* (2020), 1023–30.

[22] Bobonis, G. J. The impact of conditional cash transfers on marriage and divorce. *Economic Development and Cultural Change 59*, 2 (2011), 281–312.

[23] Boozer, M. A., and Philipson, T. J. The impact of public testing for human immunodeficiency virus. *The Journal of Human Resources 35*, 3 (2000), 419–446.

[24] Boulton, A. J., and Williford, A. Analyzing skewed continuous outcomes with many zeros: A tutorial for social work and youth prevention science researchers. *Journal of the Society for Social Work and Research 9*, 4 (2018), 721–740.

[25] Bramoullé, Y., Djebbari, H., and Fortin, B. Peer effects in networks: A survey. *Annual Review of Economics 12*, 1 (2020), 603–629.

[26] Breza, E., Chandrasekhar, A., Golub, B., and Parvathaneni, A. Networks in economic development. *Oxford Review of Economic Policy 35*, 4 (2019), 678–721.

[27] Bryan, G., Chowdhury, S., and Mobarak, A. M. Underinvestment in a profitable technology: The case of seasonal migration in bangladesh. *Econometrica 82*, 5 (2014), 1671–1748.

[28] Bühler, C., and Kohler, H.-P. Talking about AIDS: The influence of communication networks on individual risk perceptions of HIV/AIDS infection and favored protective behaviors in South Nyanza District, Kenya. *Demographic Research S1* (2003), 397–438.

[29] Caeyers, B., and Fafchamps, M. Exclusion bias in the estimation of peer effects. Working Paper 22565, National Bureau of Economic Research, 2016.

[30] Canning, D. The economics of hiv/aids in low-income countries: The case for prevention. *Journal of Economic Perspectives 20*, 3 (2006), 121–142.

[31] Card, D. Estimating the return to schooling: Progress on some persistent econometric problems. *Econometrica 69*, 5 (2001), 1127–1160.

[32] Chandrasekhar, A. G., Larreguy, H., and Xandri, J. P. Testing models of social learning on networks: Evidence from two experiments. *Econometrica 88*, 1 (2020), 1–32.

[33] Conley, T. G., and Udry, C. R. Learning about a new technology: Pineapple in ghana. *American Economic Review 100*, 1 (2010), 35–69.

[34] Cornelissen, T., Dustmann, C., Raute, A., and Schönberg, U. From late to mte: Alternative methods for the evaluation of policy interventions. *Labour Economics 41* (2016), 47–60.

[35] Corno, L., Hildebrandt, N., and Voena, A. Age of marriage, weather shocks, and the direction of marriage payments. Working Papers 2017-055, Human Capital and Economic Opportunity Working Group, 2017.

[36] Corno, L., Hildebrandt, N., and Voena, A. Age of marriage, weather shocks, and the direction of marriage payments. *Econometrica 88*, 3 (2020), 879–915.

[37] Corno, L., and Voena, A. Selling daughters: age of marriage, income shocks and the bride price tradition. IFS Working Papers W16/08, Institute for Fiscal Studies, 2016.

[38] de Paula, Á., Shapira, G., and Todd, P. E. How beliefs about hiv status affect risky behaviors: Evidence from malawi. *Penn Institute for Economic Research (PIER) Working Paper Series* (2008).

[39] Debnam, K. J., Holt, C. L., Clark, E. M., Roth, D. L., and Southward, P. L. Relationship between religious social support and general social support with health behaviors in a national sample of african americans. *Journal of Behavioral Medicine 35* (2011), 179–189.

[40] Delavande, A., and Kohler, H.-P. Subjective expectations in the context of HIV/AIDS in Malawi. *Demographic Research 20*, 31 (2009), 817–875.

[41] Delavande, A., and Kohler, H.-P. The impact of hiv testing on subjective expectations and risky behavior in malawi. *Demography 49* (2012), 1011–1036.

[42] Delavande, A., and Kohler, H.-P. HIV/AIDS-related expectations and risky sexual behaviour in Malawi. *Review of Economic Studies 83*, 1 (2016), 118–164.

[43] Dell, M., Jones, B. F., and Olken, B. A. What do we learn from the weather? the new climate-economy literature. *Journal of Economic Literature 52*, 3 (2014), 740–98.

[44] Dickson, M. The causal effect of education on wages revisited. *Oxford Bulletin of Economics and Statistics 75*, 4 (2013), 477–498.

[45] Dupas, P. Do teenagers respond to hiv risk information? evidence from a field experiment in kenya. *American Economic Journal: Applied Economics 3*, 1 (2011), 1–34.

[46] Dupas, P. Short-run subsidies and long-run adoption of new health products: Evidence from a field experiment. *Econometrica 82*, 1 (2014), 197–228.

[47] Emina, J., Madise, N., Kuepie, M., Zulu, E., and Ye, Y. Identifying hiv most-at-risk groups in malawi for targeted interventions. a classification tree model. *BMJ open 3* (2013).

[48] Fafchamps, M. Causal effects in social networks. *Revue économique 66*, 4 (2015), 657–686.

[49] Fedor, T. M., Kohler, H.-P., and Behrman, J. R. The impact of married individuals learning hiv status in malawi: Divorce, number of sexual partners, and condom use with spouses. *Demography 52*, 1 (2015), 259–280.

[50] Ferrara, E. L. Kin groups and reciprocity: A model of credit transactions in ghana. *American Economic Review 93*, 5 (2003), 1730–1751.

[51] Fulford, S. The puzzle of marriage migration in India. Boston College Working Papers in Economics 820, Boston College Department of Economics, 2012.

[52] Garner, R. C. Safe sects? dynamic religion and aids in south africa. *The Journal of Modern African Studies 38*, 1 (2000), 41–69.

[53] Gerrard, M., Gibbons, F., and Bushman, B. Relation between perceived vulnerability to hiv and precautionary sexual behavior. *Psychological bulletin 119* (1996), 390–409.

[54] Godlonton, S., and Thornton, R. Peer effects in learning hiv results. *Journal of development economics 97* (2012), 118–129.

[55] Godlonton, S., and Thornton, R. L. Learning from others' hiv testing: Updating beliefs and responding to risk. *American Economic Review 103*, 3 (2013), 439–44.

[56] Goldberg, J., Macis, M., and Chintagunta, P. Leveraging patients' social networks to overcome tuberculosis underdetection: A field experiment in india. *SSRN Electronic Journal* (2018).

[57] Golub, B., and Jackson, M. O. How homophily affects the speed of learning and best-response dynamics. *The Quarterly Journal of Economics 127*, 3 (2012), 1287–1338.

[58] Gong, E. Hiv testing and risky sexual behaviour. *The Economic Journal 125*, 582 (2015), 32–60.

[59] Gray, C. L., and Mueller, V. Natural disasters and population mobility in bangladesh. *Proceedings of the National Academy of Sciences 109*, 16 (2012), 6000–6005.

[60] Green, E., Halperin, D., Nantulya, V., and Hogle, J. Uganda's hiv prevention success: The role of sexual behavior change and the national response. *AIDS and behavior 10* (2006), 335–346.

[61] Green, E. C. *Faith-based organizations: Contributions to HIV prevention.* United States Agency for International Development, 2003.

[62] Griliches, Z. Estimating the returns to schooling: Some econometric problems. *Econometrica 45*, 1 (1977), 1–22.

[63] Grimm, V., and Mengel, F. Experiments on belief formation in networks. *Journal of the European Economic Association 18*, 1 (2018), 49–82.

[64] Hale, G., and Brathwaite, B. Response of religious groups to hiv/aids as a sexually transmitted infection in trinidad. *BMC public health 5* (2005), 121.

[65] Hartwig, K., Kissioki, S., and Hartwig, C. Church leaders confront hiv/aids and stigma: A case study from tanzania. *Journal of Community & Applied Social Psychology 16* (2006), 492–497.

[66] Heckman, J. J., and Vytlacil, E. Structural equations, treatment effects, and econometric policy evaluation. *Econometrica 73*, 3 (2005), 669–738.

[67] Helleringer, S., and Kohler, H.-P. Social networks, perceptions of risk, and changing attitudes towards hiv/aids: New evidence from a longitudinal study using fixed-effects analysis. *Population Studies 59*, 3 (2005), 265–282.

[68] Henry, S., Schoumaker, B., and Beauchemin, C. The impact of rainfall on the first out-migration: A multilevel event-history analysis in burkina faso. *Population and Environment - POP ENVIRON 25* (2004), 423–460.

[69] Hess, S., Jaimovich, D., and Schündeln, M. Development projects and economic networks: Lessons from rural Gambia. *The Review of Economic Studies 88*, 3 (2020), 1347–1384.

[70] Hoddinott, J., and Skoufias, E. The impact of progresa on food consumption. *Economic Development and Cultural Change 53*, 1 (2004), 37–61.

[71] Hoffmann, R. Following the peers: The role of social networks for health care utilization in the philippines. Vienna Institute of Demography Working Papers 17/2017, Vienna Institute of Demography, Vienna, 2017.

[72] Iyer, S. *Religion and economic development*. Palgrave Macmillan UK, London, 2010, pp. 222–228.

[73] Jackson, M. The economics of social networks. *California Institute of Technology, Division of the Humanities and Social Sciences, Working Papers* (2005).

[74] Jackson, M. An overview of social networks and economic applications. *Handbook of Social Economics 1* (2011).

[75] Jackson, M. O. *Social and economic networks*. Princeton University Press, USA, 2008.

[76] James, J. Peer effects in free school meals: Information or stigma? *EUI MWP Working Paper* (2012).

[77] Kaplan, G. Moving back home: Insurance against labor market risk. *Journal of Political Economy 120*, 3 (2012), 446–512.

[78] Kim, J. The effect of peers on hiv infection expectations among malawian adolescents: Using an instrumental variables/school fixed effect approach. *Social Science & Medicine 152* (2016), 61–69.

[79] Kinnan, C., Samphantharak, K., Townsend, R., and Vera-Cossio, D. Insurance and propagation in village networks. Discussion Paper 115, Puey Ungphakorn Institute for Economic Research, 2019.

[80] Kinnan, C., and Townsend, R. Kinship and financial networks, formal financial access, and risk reduction. *American Economic Review 102*, 3 (2012), 289–93.

[81] Kohler, H.-P., Behrman, J. R., and Watkins, S. C. Social networks and hiv/aids risk perceptions. *Demography 44*, 1 (2007), 1–33.

[82] Kohler, H.-P., Watkins, S. C., Behrman, J. R., Anglewicz, P., Kohler, I. V., Thornton, R. L., Mkandawire, J., Honde, H., Hawara, A., Chilima, B., Bandawe, C., Mwapasa, V., Fleming, P., and Kalilani-Phiri, L. Cohort profile: The Malawi Longitudinal Study of Families and Health (MLSFH). *International Journal of Epidemiology 44*, 2 (2014), 394–404.

[83] Langlie, J. K. Social networks, health beliefs, and preventive health behavior. *Journal of Health and Social Behavior 18*, 3 (1977), 244–260.

[84] Linnemayr, S., and MacCarthy, S. Using behavioral economics to promote hiv prevention for key populations. *Journal of AIDS & Clinical Research 9* (2018).

[85] Liu, X., Patacchini, E., and Zenou, Y. Endogenous peer effects: Local aggregate or local average? *Journal of Economic Behavior & Organization 103* (2014).

[86] Loutfi David, Andersson Neil, L. S. Can social network analysis help to include marginalised young women in structural support programmes in Botswana? A mixed methods study. *International Journal for Equity in Health 18*, 1 (2019), 1–11.

[87] Lucas, R. E. B., and Stark, O. Motivations to remit: Evidence from botswana. *Journal of Political Economy 93*, 5 (1985), 901–918.

[88] Malde, B. Socially close and distant connections in risk sharing. In *Draft* (2016).

[89] Manski, C. F. Identification of endogenous social effects: The reflection problem. *The Review of Economic Studies 60*, 3 (1993), 531–542.

[90] Marsden, P. V., and Campbell, K. E. Measuring tie strength. *Social Forces 63*, 2 (1984), 482–501.

[91] Maselko Joanna, Hughes Cayce, C. R. Religious social capital: its measurement and utility in the study of the social determinants of health. *Social science & medicine 73*, 5 (2011), 759–767.

[92] Maxwell, S. A radical rethinking of the way to fight global poverty. *Journal of International Development 25*, 7 (2013), 1026–1028.

[93] Morten, M. Temporary migration and endogenous risk sharing in village india. *Journal of Political Economy 127*, 1 (2019), 1–46.

[94] Munshi, K. Networks in the modern economy: Mexican migrants in the u.s. labor market. *The Quarterly Journal of Economics 118*, 2 (2003), 549–599.

[95] Musgrove, P. A. *Health economics in development*. The World Bank, 2003.

[96] Ngatia, M. Social interactions and individual reproductive decisions. In *Draft* (2011).

[97] Novignon, J., Novignon, N., Aryeetey, G., and Nonvignon, J. Hiv/aids-related stigma and hiv test uptake in ghana: evidence from the 2008 demographic and health survey. *Etude de la Population Africaine 28* (2014), 1362–1379.

[98] Olsen, M. K., and Schafer, J. L. A two-part random-effects model for semi-continuous longitudinal data. *Journal of the American Statistical Association 96*, 454 (2001), 730–745.

[99] Oluduro, O. The role of religious leaders in curbing the spread of HIV/AIDS in Nigeria. *Potchefstroom Electronic Law Journal (PELJ) 13* (2010), 208–236.

[100] Oreopoulos, P. Estimating average and local average treatment effects of education when compulsory schooling laws really matter. *American Economic Review 96*, 1 (2006), 152–175.

[101] Oster, E. Hiv and sexual behavior change: Why not africa? *Journal of Health Economics 31*, 1 (2012), 35–49.

[102] Paiva, V., Garcia, J., Rios, L., Santos, A., Terto, V., and Munõz-Laboy, M. Religious communities and hiv prevention: An intervention study using a human rights-based approach. *Global Public Health 5*, 3 (2010), 280–294.

[103] Pauw, K., Dorosh, P. A., and Mazunda, J. Exchange rate policy and devaluation in Malawi. IFPRI discussion papers 1253, International Food Policy Research Institute (IFPRI), 2013.

[104] Puente, G., Perez, F., and Gitter, R. The effect of rainfall on migration from mexico to the u.s. *International Migration Review 50* (2015).

[105] Rosenstock, I. M. Historical origins of the health belief model. *Health Education Monographs 2*, 4 (1974), 328–335.

[106] Rosenzweig, M. R., and Stark, O. Consumption smoothing, migration, and marriage: Evidence from rural india. *Journal of Political Economy 97*, 4 (1989), 905–926.

[107] Rosenzweig, M. R., and Wolpin, K. I. Parental and public transfers to young women and their children. *The American Economic Review 84*, 5 (1994), 1195–1212.

[108] Salcedo, A., Schoellman, T., and Tertilt, M. Families as roommates: Changes in u.s. household size from 1850 to 2000. *Quantitative Economics 3*, 1 (2012), 133–175.

[109] Schaefer, R., Thomas, R., Maswera, R., Kadzura, N., Nyamukapa, C., and Gregson, S. Relationships between changes in hiv risk perception and condom use in east zimbabwe 2003-2013: Population-based longitudinal analyses. *BMC Public Health 20* (2020).

[110] Schneider, J., Zhou, A., and Laumann, E. A new hiv prevention network approach: Sociometric peer change agent selection. *Social science & medicine 125* (2014).

[111] Skoufias, E. Progresa and its impacts on the welfare of rural households in mexico. Research Report 139, International Food Policy Research Institute, 2005.

[112] Smith, K. P. Why are they worried? Concern about AIDS in rural Malawi. *Demographic Research S1* (2003), 279–318.

[113] Smith, K. P., and Watkins, S. C. Perceptions of risk and strategies for prevention: Responses to HIV/AIDS in rural Malawi. PGDA Working Papers 0305, Program on the Global Demography of Aging, 2005.

[114] Smith, R., and Baker, M. At the edge? hiv stigma and centrality in a community's social network in namibia. *AIDS and Behavior 16*, 3 (2012), 525–534.

[115] Staveteig Ford, S., Wang, S., Head, S., Bradley, S., and Nybro, E. Demographic patterns of hiv testing uptake in sub-saharan africa. *DHS Comparative Report no* (2013).

[116] Stock, J., and Yogo, M. *Testing for weak instruments in linear IV regression.* Cambridge University Press, New York, 2005, pp. 80–108.

[117] Surur, F., and Kaba, M. The role of religious leaders in hiv/aids prevention, control, and patient care and support: A pilot project in jimma zone. *Northeast African Studies 7*, 2 (2000), 59–79.

[118] Thornton, R. L. The demand for, and impact of, learning hiv status. *American Economic Review 98*, 5 (2008), 1829–63.

[119] Trinitapoli, J. Religious responses to aids in sub-saharan africa: An examination of religious congregations in rural malawi. *Review of Religious Research 47*, 3 (2006), 253–270.

[120] Trinitapoli, J. Religious teachings and influences on the ABCs of HIV prevention in Malawi. *Social Science & Medicine 69*, 2 (2009), 199–209.

[121] Trinitapoli, J. The aids-related activities of religious leaders in malawi. *Global Public Health 6*, 1 (2011), 41–55.

[122] UNAIDS. *Knowledge is Power.* UNAIDS, Geneva, 2018.

[123] UNAIDS. *UNAIDS Data 2019*. UNAIDS, Geneva, 2019.

[124] UNDP. *Human Development Report 2019*. UNDP, New York, 2019.

[125] Valente, T. W. Network interventions. *Science 337*, 6090 (2012), 49–53.

[126] Venables, E. Unimagined community: sex, networks and aids in uganda and south africa. *Africa 81*, 4 (2011), 675–676.

[127] Ward, H. Prevention strategies for sexually transmitted infections: importance of sexual network structure and epidemic phase. *Sexually Transmitted Infections 83*, suppl 1 (2007), i43–i49.

[128] Watkins, S. C. Navigating the aids epidemic in rural malawi. *Population and Development Review 30*, 4 (2004), 673–705.

[129] Wiginton, J. M., King, E. J., and Fuller, A. O. 'we can act different from what we used to': Findings from experiences of religious leader participants in an hiv-prevention intervention in zambia. *Global Public Health 14*, 5 (2019), 636–648.

[130] Witoelar, F. Risk sharing within the extended family: Evidence from the indonesia family life survey. *Economic Development and Cultural Change 62*, 1 (2013), 65–94.

[131] Ye, Q., Wu, B., and Wang, B. Distance distribution and average shortest path length estimation in real-world networks. In *Advanced Data Mining and Applications* (Berlin, Heidelberg, 2010), Springer Berlin Heidelberg, pp. 322–333.