Social Protection, Health Risk, and Household Welfare in Zambia

Peter Hangoma

Thesis for the degree of philosophiae doctor (PhD)
at the University of Bergen

2017

Date of defence: 3 May 2017

Scientific environment

Centre for International Health

Department of Global Public Health and Primary Care

University of Bergen

Bergen, Norway

Department of Economics,

University of Bergen,

Bergen, Norway

Contents

List of Figures

Acknowledgements

First, I want to most graciously thank my supervisors, Bjarne Robberstad from the Centre for International Health (CIH) and Arild Aakvik from the Department of Economics for the guidance, support, feedback and critiques at various stages of my PhD. Theirs was a rear combination which any graduate student would be lucky to have. Special appreciation go to the Central Statistical Office (CSO) of Zambia and MEASURE for providing the data used in this thesis.

I also wish to thank the two people who introduced me to health, Felix Masiye and Charles Michelo. When I moved to the health, from monetary economics, I did not imagine it could turn out this meaningful, interesting, challenging, and adventurous. I owe all this to Felix and Charles as well as friends and mentors at the Institute for Health Metrics and Evaluation (IHME). Sincere gratitude also goes out to the Norwegian School of Economics (NHH) and the Swiss School of Public Health for granting me the opportunity to be a visiting PhD student.

My heartfelt appreciations are extended to the National Institute of Health (NIH), Forgarty International Center, particularly Sten Vermund , then at Vanderbilt Institute for Global Health (VIGH), and his great team, including Doug Heimburger and Ben Chi, for awarding me the scholarship to study in Bergen. In this vein, special mention goes out to Holly Cassell who was always in contact to ensure that I did not have any logistical challenges. I also thank my employer, the University of Zambia, for giving me the chance to pursue my studies. Support from Norwegian Partnership Programme for Global Academic Cooperation (NOPART) is much appreciated.

To the University of Bergen, CIH which housed and provided me with all manner of support during my PhD work, I say thank you. In particular, I thank Bente Moen and Bjarne Robberstad for supporting and involving me in many activities at CIH. Appreciations are also extended to Ingvild Hope, Gunhild Koldal, Borgny Lavik, Solfrid Vikoren, and Øyvind Mørkedal, who were in hand to provide administrative assistance. And I shall always trea-

sure the memories of running, hiking, or mountain biking with Ingunn Engebretsen, Karen Moland, Arild Aakvik, Mengs Tewelde, Lars Fadness, Linda Forshaw, Mengistu Bekele, Mehma Pannu, Bente Moen, Melf Jakob, Beatrice Odongkara, Bjarne Robberstad, Alemayehu Hailu, Jo Even Warpe, Jovita Kakia, Melissa Jørstad, Angela Akol, Pascal Ruggajo,Akeza Awealom, Robin Cole, Mwila Lundamo, Jorn Blume, Alseme Sanu, Simon Kibira, Hanne Keyser, and Sudha Basnet. I also wish to thank all members of the health economics meetings and the bigger Global Health Priorities research group for the thoughtful engagements we had.

My wife, Kapema, is indescribable and her stay with me in Bergen brightened my life. To my kids, Peter Libazisa, Nthanda, and Magnus Jabulani, the times we spent playing blew away all the stress that could have otherwise crippled me. I am thankful for the encouragement I have received from my mother, Opah Moyo, my sisters, Linda, Charlen, Charllotte, Matilda and Corette, as well as my grandfather, Joseph Hangoma.

All the days we would talk about happiness with Ana Hammer and Maija Rankinen shall remain priceless. I also appreciate the time spent with friends, Richard Banda, Nkomba Kayeyi, Jobiba Chinkhumba, Ulrikke Voltersvik, Elham Nikpey, Linda Kampata, and Amani Thomas. I am also thankful to the Zambia–Norway friendship association in Bergen, they are real family.

List of papers

1. *Peter Hangoma, Arild Aakvik, and Bjarne Robberstad*

   Health shocks and household welfare in Zambia: An assessment of changing risk .

2. *Peter Hangoma, Bjarne Robberstad, and Arild Aakvik*

   Does Free Public Health Care Increase Utilization and Reduce Spending? Heterogeneity and Long Term Effects.

3. *Peter Hangoma, Arild Aakvik, and Bjarne Robberstad*

   Explaining changes in child health inequality in the run up to the 2015 Millennium Development Goals (MDGs): The case of Zambia.

Summary

Households in sub-Saharan Africa face substantial health risk. This threatens their welfare and predisposes them to poverty. Despite the high risk environment, they have little or no access to social protection–a set of programs that aims to reduce health risk and provides insurance against its effects, key of which are reductions in labor income and increases in household health expenditure. In childhood, health risk may have additional effects; it lowers cognitive abilities as well as educational attainment and these effects persist in adulthood, working to permanently lower lifetime economic outcomes. Yet still, children from poorer backgrounds face a disproportionately larger share of childhood health risk.

In this thesis, I examined the extent to which households are protected from the welfare effects of health shocks (illness and injury) in Zambia. I also evaluated some social protection policies focused at the general population and specific groups such as children and individuals from low socioeconomic background. This was achieved in three sub-studies, each of which forms a separate paper. The first one assessed the effect of health shocks on household consumption, income, and health spending, as well as the extent to which households use borrowing and selling assets as coping strategies in the absence of complete social protection systems, during and after structural adjustment reforms (SAPs). Using data from four waves of the living conditions monitoring survey (LCMS) in the period 1996–2006, it was found that health shocks were associated with reduced consumption both during and after structural reforms. Although health shocks were substantially associated with reduced labor income in both periods, the effect on health spending was much greater after the structural reforms. Middle income households were especially vulnerable. To cope with this risk, household employed informal borrowing and selling assets as self insurance mechanisms.

In the second paper, the short and long term effects of an important social protection policy–the user fee removal–on medical spending and overall utilization of health services was evaluated. Heterogeneity in utilization response was also examined. Results show that

the policy increased overall utilization of health services in the short term and these effects were sustained in the long term. Apart from increasing overall utilization, the policy also led to shifting of use from private to public health services. The greatest increase in utilization of health services occurred among individuals whose household heads were either unemployed or had no education. Further, although the policy reduced the proportion of individuals incurring any spending, overall health expenditure was not affected in any significant way.

Third, and finally, the last paper investigated the determinants of childhood health risk, specifically stunting and fever, between 2007–2014, a period of massive scale up of child health interventions as countries braced themselves to meet the 2015 target of the Millennium Development Goal on child health. It assessed whether or not the concentration of health risk among children from poorer households reduced. Importantly, the factors or determinants that could have been driving these changes were investigated. It was found that although the prevalence of stunting in the general population and in all quartiles, except the poorest, reduced, inequality increased significantly. The determinants that contributed the most to the increase in inequality of stunting were maternal height and weight, household wealth, birth order, place of birth (home or facility), breastfeeding duration and maternal education. Socioeconomic inequality in fever also increased and incidence of fever did not fall. The determinants that contributed to the increase in the inequality of fever were household wealth, maternal education, birth order, and duration of breast feeding.

I conclude that scaling up social protection programs that aim at providing insurance against health risk would improve household welfare, especially if coverage does not only focus on the poorest but also middle income households, who are found to be most vulnerable. Regarding user fee removal, although this policy may reduce health risk, since it increased utilization of health services, especially among individuals from low socioeconomic backgrounds, it was not successful in reducing health expenditure risk. Other social protection programs need to be considered if there has to be improvement in health spending insurance. In the same vein, childhood health risk became more concentrated among children from poorer households despite the massive scale up in child health interventions. Reducing inequality in the determinants of childhood health such as facility deliveries, wealth, education, nutrition, etc is key to reducing inequalities in childhood health risk. If inequalities of determinants are not eliminated, increasing their coverage may not reduce child health inequality, and may, in fact, increase it.

Chapter 1

Introduction

1.1 Background

Consumption of goods and services (such as nutritious food, education, housing, clean water, etc.) is an important indicator of household welfare. However, the ability of households to maintain adequate levels of consumption may be threatened by both covariate risk (e.g, drought, earthquakes, floods etc–that is, risk which, when realized, affects all households at once) and idiosyncratic risk (e.g health deterioration, job loss, crop failure etc–risk whose realization affects only one or a few households at any given time). A shock is said to have occurred when risk is realized, resulting in welfare losses. By definition, idiosyncratic risk, unlike covariate risk, can be insured more easily through appropriate social protection programs [1]. Moreover, relevant social protection interventions can be used to reduce or minimize idiosyncratic risk, a sheer contrast with covariate risk which mainly relates to events beyond human control.

Yet still, one type of idiosyncratic risk, namely, health (sickness, injury and death), poses one of the most serious threats to household welfare [2, 3]. Unlike other idiosyncratic shocks such as job losses which may only lower labor income, health shocks may lead to both reductions in labor income [4] and substantial increases in health expenditure [5, 6], ultimately lowering household consumption, in the absence of social protection. To maintain stable consumption, especially if already at subsistence level, households may further employ a number of informal self insurance or coping strategies such as selling assets, taking children out of school and borrowing [7]. Informal coping strategies may have long term welfare consequences [8], including keeping households in perpetual poverty and contributing to the inter-

generational transmission of poverty. In childhood, health shocks may have additional short and long term consequences. Parents may need to take time off work in order to care for their sick children, resulting in lower labor income earnings in the short to medium term. In the long term, health shocks may lower a child's overall health status, cognitive abilities, as well as educational attainment, and these effects are not only persistent over time, but they permanently lower lifetime incomes and other economic outcomes [9–12]. This implies that a child's health shocks may have welfare consequences on both her parents' household and her own future household.

Put together, the short and long term welfare losses that health shocks induce may be substantial, driving households into poverty or keeping them in perpetual destitution. This is especially true if they face high vulnerability, the level of which depends on the magnitude of health risk, availability of social protection (in terms of cost of insurance/social assistance, coverage, and adequacy of benefits), initial consumption levels (poverty status), and the responsiveness of medical spending as well as labor income to health shocks.[1] Households are not vulnerable, and will not suffer welfare loss, if they do not face any health risk. For a given level of health risk, a household with lower consumption levels (living at bare subsistence) is more likely to be driven into poverty, and hence more vulnerable. Similarly, households are more vulnerable when they face a high cost of insurance or social protection, e.g, high interest rates from self insurance strategies such as informal borrowing. In the same vein, households with higher increases in health spending and reductions in earned income are more vulnerable because they may experience larger drops in consumption as a result. But even when health expenditure and labor income risks are high, welfare may not be affected, and households are thus not vulnerable, if social protection systems provide transfers that offset the downward

[1] One of the most commonly used benchmark for assessing household welfare is a nationally or internationally defined consumption threshold, known as the poverty line. In this case, a household is said to be vulnerable if its consumption has a high likelihood of falling below a poverty line when hit by a shock [13], given all available insurance mechanisms, both public and private (formal and self/informal insurance). Formally, suppose a household with consumption level of $c > 0$ faces a health shock, h, with non-zero probability, p. It insures the health shock with an amount, $d \geq 0$. This insurance covers both possible health spending increases, m, and labor income losses, y, resulting from a shock. if the household suffers a loss of s as a result of the health shock, it receives an insurance transfer of t, where $0 \leq t \leq s$ and $s = s(m,y)$. If $t = s$, a household is fully insured. Given a consumption level, z, which defines a poverty threshold, a household is said to be vulnerable, $V(\cdot)$, if;

$$V(\cdot) = p \times (c - d - s(m,y) + t) < z \tag{1.1}$$

In words, a household is vulnerable if the expected consumption after the shock is lower than the poverty threshold, z. From this formulation, vulnerability depends on: 1.) the probability of experiencing the health shock, 2.) initial consumption level, 3.) the cost of insurance or social protection, d, 4.) the transfer amount to protect households from the effect of shocks, t, and 5.) the size of the consumption drop, $s(\cdot)$, which in turn is dependent on health spending, and labor income, both of which are determined by the severity/size of the shock.

pressure on consumption.

It turns out that these determinants of vulnerability reinforce each other; households with a higher concentration of health risk have lower levels of consumption (poorer), chiefly rely on costly informal self insurance mechanisms, and have the least access to formal social protection. The concentration of health risk among poor households is seen across various measures of socioeconomic status such as wealth [14], education [15, 16], and occupation [17–20]. This disproportionate concentration of health risk on individuals of lower socioeconomic backgrounds is termed socioeconomic health inequalities.

Of particular concern is the existence of large socioeconomic inequalities in child health [21, 22]. While some inequalities in adult health across the socioeconomic distribution may be attributed to differences in behaviors such as eating habits, smoking, exercise, etc., [23], a substantial portion of inequalities in childhood health are determined by parental socioeconomic background and health behaviors as well as the their environment [9]. Such inequalities raise concerns because children have no control over these circumstances. Hence, the bulk of socioeconomic inequalities in early childhood are regarded as unjust and unfair [14]. There is need, as such, to not only reduce health risk in children but also wipe out the unfair concentration of this risk among children from poor backgrounds.

The question arises; how can the general population and specific subgroups be better protected from health risk and its effects? From our discussion of the determinants of vulnerability, it is clear that policy may pursue three possible goals; (1) eliminate or reduce health risk, (2) provide adequate insurance against the effect of health shocks; increased health spending and reduced labor income, and (3) increase levels of consumption in order to improve household resilience to shocks. Social protection constitutes an important set of programs through which these goals can be achieved. It reduces vulnerability in the general population and in specific populations subgroups, through a system of transfers in cash or in kind [24]. These programs can be classified as either contributory or non-contributory. For contributory programs, individuals contribute resources to a central fund, or provide labor, before they can be eligible for benefits. On the other hand, eligibility for non-contributory social protection, sometimes called social assistance, is based on assessed levels of vulnerability and borders on concerns of reducing inequality. Generally, they do not require contributions or previous labor market attachment from eligible recipients. The important thing for health shocks is that both contributory and non-contributory social protection programs are used to achieve one or more

of the three goals above.

Interventions in the health sector typically focus on the first goal–reducing health risk. However, other social sectors interventions can also reduce health risk, for example, providing free clean water, and improving sanitation in poor areas can prevent childhood diarrhea. Food support programs for vulnerable groups (very poor or high health risk) can reduce malnutrition and improve child health. Similarly, the second goal–providing insurance against the effects of health shocks–is multi-sectoral. Protecting household from large medical spending (health expenditure risk) may involve a robust health sector that either freely provides quality services or manages a universal health insurance scheme that is sufficiently funded. Unlike health expenditure risk however, providing insurance against income risk is beyond the health sector. It requires well developed programs in other sectors e.g, social insurance (e.g pension, disability, maternity and sick leave benefits) and labor protection programs. For the third goal–increasing levels of consumption–robust economic and anti-poverty policies such as family cash benefits may be required. Vulnerability is minimized if social protection interventions addressing these three goals are well developed throughout the life of household members. As such, the life cycle provides the best lens of visualizing the possible social protection programs available to households, and why they are needed at each stage.

The first stage in the life cycle can be viewed to be the period when a child is still developing in her mother's womb (in-utero). As is the case in early childhood, health shocks in-utero can have long term consequences. Social protection programs at this stage thus seek to reduce health risk of a developing child. Most interventions are in the health sector and aim at ensuring quality maternal health care and nutrition so as to promote healthy development of the child, and minimize in-utero insults.

In the immediate period following birth, children need care as they adjust to the new environment. A number of social protection programs are implemented, one of which is maternity protection. This is an important avenue of reducing childhood health risk. Specifically, maternity leave may give a mother enough time to care for the new born and identify possible health dangers in the first critical months of childhood. Although maternity leave is important in itself, making sure that it is a paid one (mother still drawing a salary) may prevent premature return to work. Other benefits that seek to address any combination of the three goals of social protection we have identified, namely, reducing health risk, insuring health risk, and increasing consumption, may be available in childhood. They include free health care, cash transfers

to families with children under certain ages, school child feeding programs, etc. Most of these benefits are from non-contributory social protection systems and are designed to weaken the dependency of child health and later life outcomes on parental circumstances.

After childhood, when an individual reaches a legal labor force age, they either get formally employed or not. If they do, various social insurance and labor benefits are available which provide both income and medical spending insurance as well as reducing health risk. If they suffer a health shock, they may be protected from the risk of incurring substantial medical spending through access to social health insurance schemes or employer provided private health insurance. For income risk, depending on the magnitude and duration of a health shock, individuals may have access to sick leave, short term disability benefits, and long term disability benefits. Additionally, an individual may be entitled to unemployment benefits if they are laid off before reaching retirement age. Such benefits, though not directly related to health shocks reduce vulnerability by improving household resilience to shocks. At retirement, individuals may be entitled to cash benefits in the form of pension. This is important because old age is normally associated with increased health risks.

Adults who either have been unemployed or are working in the informal sector may not have access to income and medical spending insurance that comes with contributory social protection (disability benefits, pension, maternity benefits, health insurance etc). As such, their only recourse is social assistance programs such as social cash transfer, food stamps, etc. By increasing consumption and making it independent of health shocks, these programs may improve resilience and reduce household vulnerability. Such individuals may also benefit from free health services or universal health insurance provided to the entire population, regardless of employment status. This may reduce health risk–improve health status–in addition to minimizing health expenditure risk. However, these individuals, by virtue of being unemployed or working in the informal sector, may still face substantial income risk. This is particularly important for countries in sub-Saharan Africa and south Asia where informal sector shares of the labor market is as high as 90% [25].

The foregoing life cycle perspective shows that social protection may protect households from health risk and improve welfare. As such, it has been heralded as a vehicle for achieving a number of Sustainable Development Goals (SDGs), especially the three relating to poverty, health, and inequality. The importance of social protection is also reflected in the Universal declaration of human rights which says that “Everyone, as a member of society, has the right to

social security..." [26]. Despite its importance, about three quarter of the world population do not have access to comprehensive social protection, and the majority of these live in poverty [27]. Informal self insurance (coping) strategies become key in households' quest to protect themselves from the effect of health shocks [28]. Although, such strategies may be viewed as substitutes for social protection schemes (formal private or public insurance and social assistance), these strategies may deplete the asset base of poor households, lead to long term poverty, and threaten household capacity to cope with future shocks [29]. Hence, a more important question in assessing vulnerability is not whether or not self coping strategies are effective in insuring households against health shocks at any given time, but the extent to which households rely on such potentially inefficient strategies when hit by health shocks.

Sub-Saharan Africa presents a particularly strong case of vulnerability; being the region with the largest health risk, highest poverty rates, and yet having the lowest coverage of social protection. Although the development challenges are complex, the high levels of vulnerability have been partly attributed to the tightening structural adjustment reforms, implemented between the 1980s to late 1990s [30–32]. Among other things, the reforms led to introduction of user fees for health services and associated reductions in health service utilization[33], reduction in social sector funding [34], and shutting down of companies which resulted in job losses [35]. To reduce the extent of vulnerability, most countries started strengthening or implementing poverty reduction and social protection programs in the early 2000s under Poverty Reduction strategy papers (PRSPs), programs that came with debt relief and increased external financial support [36]. Health interventions were substantially scaled up and there was a resolve to remove user fees as a way of increasing utilization of health services, improving health status as well as providing financial protection against health expenditure risk [37]. Assessing the extent of vulnerability–that is, how health shocks impact on welfare, the magnitude of income and health expenditure risk, how households cope with health risk and how some social protection policies have performed in terms of reducing this risk as well as providing health spending insurance–is important in informing policy on the appropriate design of social protection programs.

Using the case of Zambia, I pursue three tasks. First, I question whether, taken together, all social protection systems sufficiently protect households from the welfare damaging effects of health shocks. I examine the effect of health shocks on household consumption and the paths through which they may be transmitted to consumption. Is it through reduced labor income or

increased health expenditure? In other words, I attempt to quantify the magnitude of income and health expenditure risk. I also assess the extent to which household employ some self insurance strategies when faced with health shocks.

Second, I evaluate the effect of one important social protection policy, the removal of user fees in public health facilities. The goal of the policy was to increase access to health services by reducing financial barriers. Increasing utilization of health services may improve health status, and reduce health risk. Additionally, removing user fees may lower household medical spending, and hence, lower health expenditure risk. We ask whether the policy was successful in increasing overall utilization of health service and reducing health expenditure. Were these effects sustained in the long term? How did health seeking behavior of individuals from different socioeconomic backgrounds respond to the removal of user fees?

Third, I assess changes in childhood health risk, specifically stunting and fever incidence, during the period of massive scale-up of child health interventions. Was there inequality in childhood health risk? or stated differently, was there a higher concentration of health risk among children from poor backgrounds? Did this inequality change? What factors were driving these changes?

1.2 Vulnerability in sub-Saharan Africa

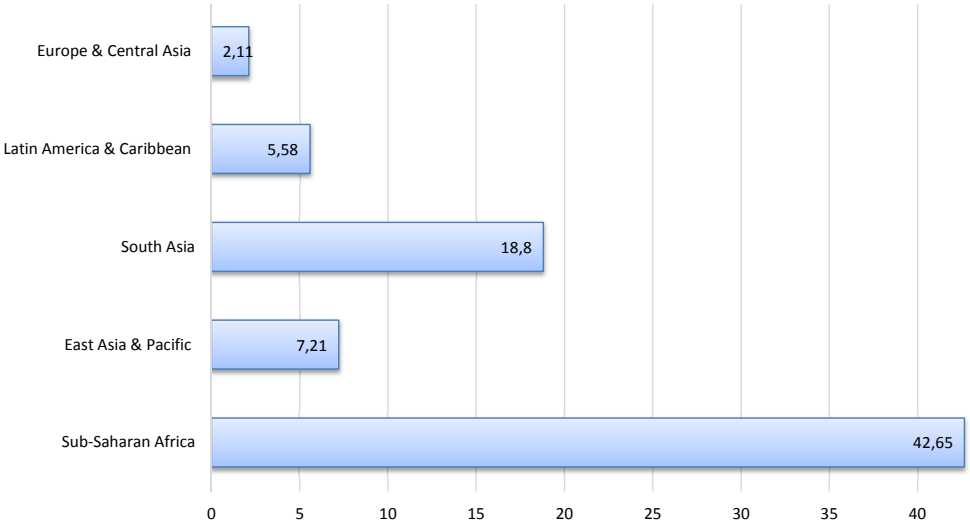
As mentioned earlier, the extent of vulnerability depends on initial consumption levels or poverty status of a household, the size of health risk, and the extent to which social protection programs protect or insure households against income and health expenditure risk. Households in sub-Saharan Africa are more vulnerable than those of other regions; they have higher levels of extreme poverty, larger health risk and lower social protection. By discussing each of these aspects, this section provides an overview of the extent of vulnerability in sub-Saharan Africa relative to other regions of the world.

1.2.1 Poverty

Vulnerability in sub-Saharan Africa is underscored by the size of the population that is extremely poor–defined as living on less than $1.90 per day. Extreme poverty is a condition of severe human material deprivation. It reflects substantial challenges in meeting basic needs such as food, housing, safe drinking water, sanitation facilities, health care, education and in-

formation. According to the latest estimates from the World Bank [38], the percentage of people who live in extreme poverty in sub-Saharan Africa is more than twice that of south Asia, and six times that of east Asia and the pacific region (Figure 1.1). Although the average poverty headcount is 40% in sub-Saharan Africa, it ranges from as low as 8% in Gabon to as high as 82% in Madagascar [38].

Figure 1.1: Percentage of the population living on less than $1.90 per day in 2012 (at 2011 PPP)**



**Authors own computation based on data from the World Bank [38]

Even within countries in sub-Saharan Africa, substantial inequalities exist. Taking an example of the two most unequal countries in Africa, the consumption share of the poorest 60% is only 15.2% in South Africa and 21% in Zambia [39]. Large income inequalities may suggests substantial inequalities in access to services such as quality health care, education, clean water, and housing.

1.2.2 Health Risk

In addition to high poverty levels and socioeconomic inequalities, sub-Saharan Africa faces a disproportionately larger share of health risk than any other region in the world. It has the highest mortality rates and the lowest life expectancy, with adult mortality rate (probability of dying between 15 and 60) being more than twice the global average in 2013 (306 vs 152

per 1000 population) [40]. Although sub-Saharan Africa population is 13% of the world total, it accounts for 70% of people living with HIV [41], 88% of all malaria cases [42], 66% of maternal deaths[43], and 48% of all child deaths [44]. It has a high level of childhood under nutrition with more than 35% of children having an inadequate height for their age (stunted) [45]. Sub-Saharan Africa is the only region in the world facing persistently high levels of under-nutrition. For example, it is the only region were number of stunted children rose during the period 1990–2014 [45]..

Apart from disease, sub-saharan Africa is experiencing a rising burden of injuries and non-communicable diseases. The increase in injuries is mostly driven by the increase in road traffic accidents.The regions has the highest burden of injuries for individuals aged 0 to 49 [46]. [2] In 2012, for example, injuries caused more deaths in Africa than any other region [47].

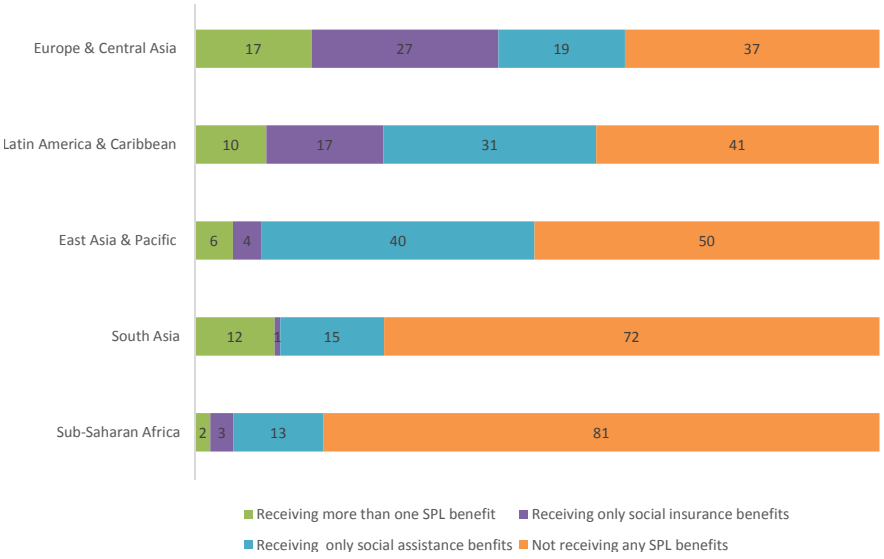
Within sub-Saharan Africa, there is substantial variation in health risk across different measures of socioeconomic status. Both mortality and morbidity is concentrated on individuals with lower income, lower educational attainment, whose occupations are involve manual labor, and who live in poorer neighborhoods.

1.2.3 Social Protection

Sub-Saharan Africa has the lowest coverage of social protection of any form, with 81% of the population not having access to social protection (Figure 1.2) [48]. Because of the large share of informal employment, only 5% of the population receive social insurance or multiple social protection and labor (SPL) benefits. In addition to the low coverage of social insurance and labor programs, social assistance coverage is also very low. East Asia and the Pacific as well as south Asia also face low coverage of social insurance and labor programs, although relatively higher than sub-Saharan Africa. The higher coverage for these regions stems from larger coverages of social assistance programs.

[2] measured in disability adjusted life years (DALYs).The DALY quantifies the total health loss for a population by combining a mortality measures (years of life lost (YLL)) and a morbidity measure (years of life lived in disability (YLD)).

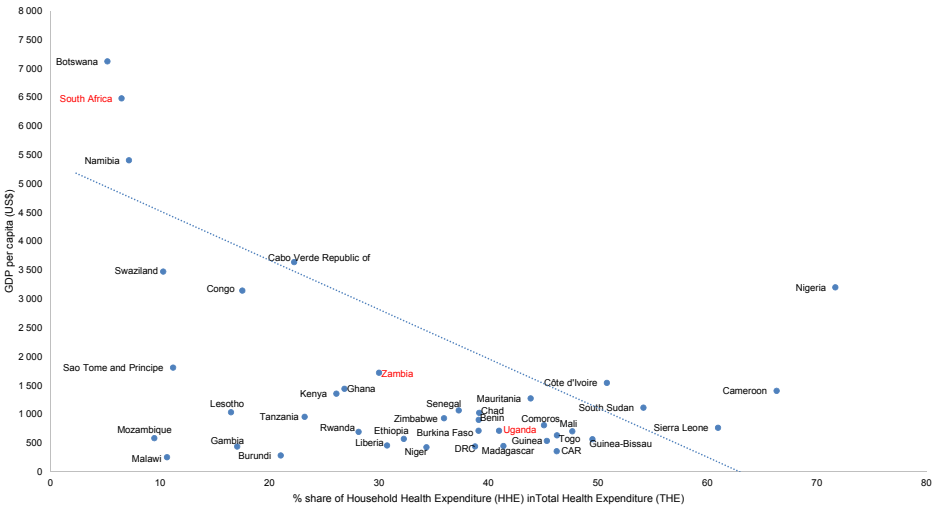
Figure 1.2: Proportion of household with access to various forms of social protection.**



**Authors own computation based on data from the World Bank [48]

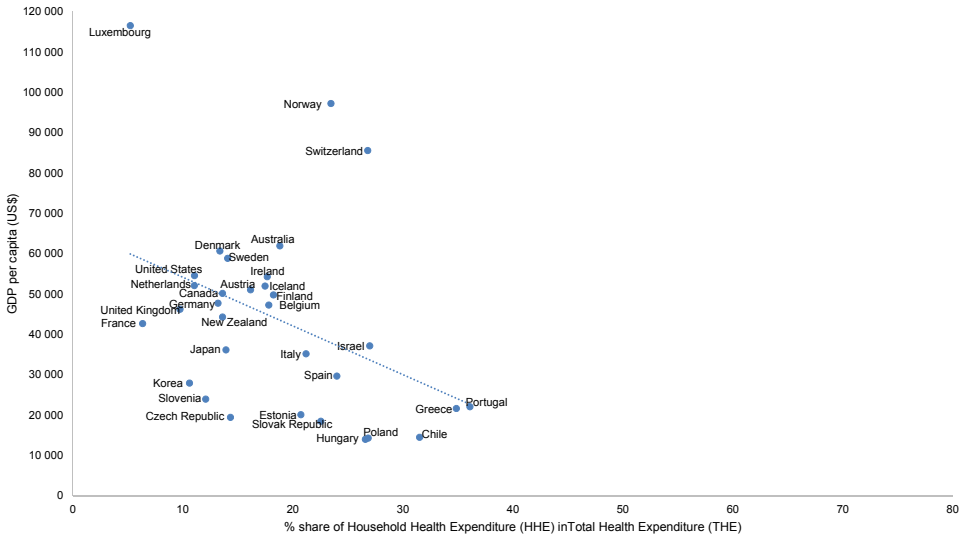
The limited coverage of social protection in sub-Saharan Africa implies not only high income risk but also health spending risk when individuals are hit by shocks. In general, household in sub-Saharan Africa face high health expenditure risk compared to household in developed countries. The risk of impoverishment is higher in countries where household share in Total Health Expenditure (THE) exceeds 15–20% [49]. In most sub-Saharan Africa countries, this share is more than 30% (Figure 1.3). Although households in OECD countries are wealthier, with median income of $38,000 dollars compared $990 in sub-Saharan African, their contribution to THE is less than 20% in most countries (Figure 1.4). The two figure also review that even within each group of countries, lower income countries are associated with higher contribution of households to THE.

Figure 1.3: GDP per capita against share of Household Health Expenditure (HHE) in sub-Saharan Africa countries, 2014**



**Authors own computation based on data from the World Health Organization [50]

Figure 1.4: GDP per capita against share of Household Health Expenditure (HHE) in OECD countries, 2014**



**Authors own computation based on data from the World Health Organization [50]

The high contribution of households to total health spending in sub-Saharan Africa in spite of low income levels has raised concerns that there may be substantial financial barriers

to health care access. In response, some countries have removed user fees for some or all the services with a view to providing financial protection and increasing utilization of health services. However, household health expenditure still remains high, even in countries, such as as Zambia and Uganda, that have completely removed user fee for all primary public health services (In red text in Figure 1.3).

Chapter 2

Setting of the study

2.1 Background

This study focuses on Zambia, a sub-Saharan Africa country located between latitudes 8° and 18° south of the equator, and longitudes 22° and 34° east of Greenwich. It is landlocked and covers an area of 752,610km^2–almost the size France and the UK combined. Cut across by a number of valleys and bathed by several river basins, Zambia sits on a plateau with an altitude of 1,000–1,600m above sea level.

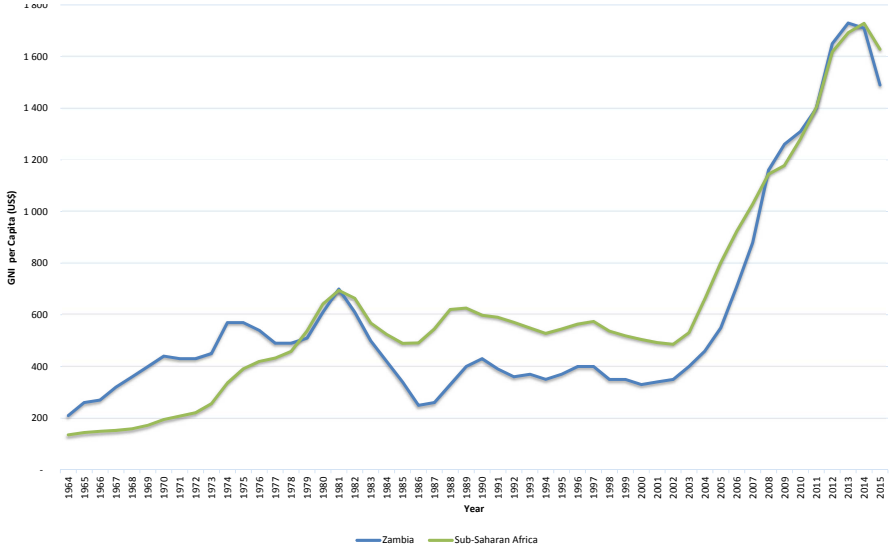
Zambia's population was estimated at 15.5 million in 2015, with an urban share of about 42% [51], making it one of the most urbanized countries in Africa. It is administratively divided into 10 provinces which are further divided into districts. Districts are the key units of decentralization where services such as health, education, etc, are delivered. The number of districts was increased from 72 in 2013 to the current 105.

From independence in 1964, Zambia enjoyed strong economic growth. It's per capita income was higher and increased at a much faster rate than the rest of sub-Saharan Africa between 1964 and 1974 (Figure 2.1, [52]). With an income per capital of about US$600 in 1974 [52], twice the sub-Saharan Africa average, Zambia was classified as a middle income country. The government owned most of the industries and imposed substantial subsidies as well as price controls on a number of essential products. In addition, education and health were freely provided. In consequence, the country enjoyed relatively high standards of living with low socioeconomic inequality.

Generally, Zambia's economic model was characterized by an almost complete dependence on copper export revenues and relatively underdeveloped agriculture sector with a large

population involved in subsistence agriculture [53]. Being one of the largest copper producers in the world, the country was able to sustain high growth as long as copper prices remained higher. Copper earnings were used to provide subsidies to government industries, most of which could not sustain themselves and it has been speculated that this might have been due to price controls, low efficiency, and other management challenges [54]. Manufacturing Industries also relied heavily on oil imports, which were financed primarily from copper export earnings. The over reliance on copper revenue constituted a recipe for Zambia's economic woes as copper prices tumbled in 1974 and export earnings fell substantially. At the same time, oil prices started rising pushing up the import bill. These two events gave rise to balance of payment and fiscal challenges which led to heavy borrowing on the international market so as to keep industries running and avoid laying off workers [54].

Figure 2.1: Income per capita between 1964 and 2015, Zambia and sub-Saharan Africa**



**Authors own computation based on data from the World Bank [52]

Even these efforts, balance of payment problems persisted, growth stalled, and Zambia had become the most indebted country in the world, relative to its gross domestic product (GDP) [54]. By the 1980s, Zambia's per capital income fell substantially lower than the rest of sub-Saharan Africa (Figure 2.1). The economic model, especially price controls, led to severe food and other commodity shortages. To curb the problem, Zambia agreed with the IMF and World Bank on a package of structural adjustment policies (SAPs) in 1983, which other countries in

sub-Saharan Africa were also pursuing. The adjustment policies included removal of subsidies and decontrol of prices. However, this worsened economic problems as prices of major commodities rose sharply. Under pressure from the population, the government suspended the SAPs and embarked on an Economic Recovery Program (ERP), reintroducing controls and other measures [55]. This program resulted in a temporal recovery of income growth from 1986 (Figure 2.1). Nonetheless, the fundamental problem still remained unresolved; price controls in the midst of high manufacturing costs resulting from high oil prices. This provided a disincentive for production and culminated into food and commodity shortages. Faced with civil unrest, the government went back to the IMF for more loans on condition that they implement a more comprehensive SAP package. However, these reforms were actually implemented by a new government which came to power in 1991.[1]

2.2 Structural Reforms and Vulnerability

A comprehensive structural adjustment program was implemented between 1991 and 2002, and involved a myriad of tightening reforms spanning the whole economy, including the health sector. The reforms included liberation, privatization, removal of subsidies, decontrol of prices, and health reforms, notable of which was the introduction of user fees in 1993 [54]. In this case, individuals seeking health services were now required to pay at the point of use. Exemptions were given to children under the age of 5 years, senior citizens over the age of 65, and poor people. Although a significant proportion of the population was exempted, mainly on account of age, most individuals that benefited were well off. For example, Masiye et al. [56] conducted an exit patient survey in five provinces and found that 53% of the patients were exempted from user fees and yet only less than 1% was on account of poverty. In addition to these reforms, the SAPs also restricted spending on social sectors such as education and health in order to keep the government budget balanced [55].

Evidence suggests that even if there were isolated cases of success in a few indicators resulting from SAPs, none of the 20 countries implementing adjustment reforms over the 1980-1999 period were able to achieve reasonable economic growth and rectify policy distortions [57]. On a micro level, it has been found that SAPs led to substantial increases in household

[1] SAPs were undertaken in many other sub-Saharan Africa countries. The only difference is that Zambia also experienced an economic crisis triggered by the sharp decline in copper prices and increase in oil prices. Thus, one can see from Figure 2.1 that from the 1980, when most countries were implementing SAPs, per capita income in Zambia and the rest of sub-Saharan Africa were tracking each other closely.

vulnerability with marked increases in poverty and inequality in Zambia [58]. By the time the reforms were completed in 2002, 67% of the population was in poverty [35]. Liberalization and privatization led to shutting down of a number of industries while those that continued operating retained a leaner work force. Massive job losses resulted [35], and the size of the informal sector increased from 20% in 1967 [59], to 83% by the end of the SAP period in 2002 [35]. The informal sector is characterized by high vulnerability because of lack of social security benefits such as paid leave, pension and other benefits.

Similarly, the introduction of user fees was associated with declines in utilization of health services [33]. Household health expenditure as a share of total health expenditure remained at over 25% in the period 1995–2002 [50]. Health indicators also worsened. According to the Demographic and Health survey [60], under five mortality rate increased from 191 to 197 per 1,000 live births between 1991 and 1996 and remained at 168 per 1,000 live births in 2001/2002.[2]. Similarly, stunting rates in children under the age of five years rose from 46% to 53% over the 1991–2002 period. Adult mortality rate also increased, from 353 to 446 per 1,000 persons between 1996 and 2001-02. HIV prevalence was estimated at 15.6% in 2001-02, one of the highest in the world. The high rate of HIV presents a special challenge by not only increasing the burden of illness and mortality but also increasing the number of children orphaned, some of whom have dropped out of school and/or are subjected to child labor.

Elsewhere in sub-Saharan Africa, SAPs were also associated with increasing food insecurity and under-nutrition, rising ill-health, and decreasing access to health care [34]. Having finished the reform process by 2002, and, cognizant of the welfare impact of SAPs, the government embarked on a program of poverty alleviation.

2.3 Social Protection and Poverty Reduction

Zambia, like other sub-Saharan Africa countries, started implemented three years plans known as poverty reduction strategy papers (PRSP) in 2002 after the SAPs were completed [35]. The overarching goal of PRSP was to reduce poverty by sustaining strong GDP growth rates that result in job creation and increased tax revenue for spending in key social sectors.

The PRSP period was associated with strong growth, not only in Zambia, but the rest of sub-Saharan Africa, with sharp increases in per capita incomes since 2002 (Figure 2.1).

[2]Note that under 5 mortality rates are averages of the 5 years proceeding the survey year while those for adult mortality are calculated for the past 7 years preceding the survey

For Zambia, this period also coincided with rising copper prices on the global markets and the cancellation of it's huge debt under the highly indebted poor countries (HIPC) initiative. However, despite registering sustained GDP growth rates of over 7% between 2006 and 2010, inequality and poverty rates have remained high. According to data from the World Bank [39], Zambia has the highest income inequality in sub-Saharan Africa, only next to South Africa, with the poorest 60% of the population consuming only 21% of national income. Poverty rates have been unresponsive, remaining at over 60% [61]. As a matter of concern, the PRSP 2002-2004, did not explicitly provide for social protection in spite of the high levels of household vulnerability.

In light of this, the country developed a social protection strategy in 2004 and co-opted it into the Fifth National Development Plan (2006–2010), as well as its successor, the Sixth National Development Plan (2011-2016). The strategy is being operationalized in the National Social Protection Policy (NSPP) which was approved in 2014. A working group, called social protection Sector Advisory Group (SP SAG), was also constituted. Its mandate is to oversee the implementation of social protection and draws on membership from government ministries, departments, as well as cooperating partners involved in implementing various components of social protection programs.

In the context of Zambia, social protection has been defined as "policies and practices that protect and promote the livelihood and welfare of the people suffering from critical levels of poverty and deprivation or vulnerability to risks and shocks". Although social protection programs are in place, severe organizational and financial challenges have inhibited the goal of reaching meaningful coverage [62]. For example, only 1.6% of the population had access to social insurance or social assistance schemes in 2010. To gain a deeper understanding of the extent of vulnerability in Zambia, we discuss its social protection system in a little more detail.

2.3.1 Contributory Social Protection

Zambia has a long history of social insurance schemes. However, their financing and administration was substantially affected by the SAPs with assets of all pension funds dropping from 10% of GDP in the 1980s to 1.6% of GDP in the late 1990s [63]. Severe challenges in paying out benefits emerged as a result, and have since persisted [64]. In the hope of improving delivery, the government revised the statutory instruments relating to pension and employment

conditions and established the National Pension Scheme Authority (NAPSA) in 2002.

NAPSA administers benefits related to retirement, disability and death of workers for all individuals joining the workforce since 2002 as well as those who were migrated from the old schemes. Contribution is compulsory for all individuals employed in the formal sector. However, benefits are only paid after a workers has made a minimum number of contributions, or has worked for a specified number of years continuously, generally 15 years. An exception is disability benefits which are given if one works for a minimum of 5 years but suffers a major illness or disability that permanently renders them incapable of working. Before being certified as incapable, the law entitles an individual to 90 days of paid sick leave on full pay and an addition 90 days on half pay if they do not recover. Thereafter, they can be discharged from employment, and would qualify for disability pension if they worked and contributed to scheme for at least 5 years. In reality however, the granting of paid sick leave is normally at the mercy of employers. In most cases, an employer, especially in the private sector, would not be willing to pay an individual for a total of 120 days (6 months).

In terms of maternity protection, only unionized female workers who have served in an organization for two years since recruitment or their last maternity leave, are provided with paid maternity leave. The leave is up to 90 days. Otherwise, just like non-unionized workers, they can take unpaid maternity leave, which can go up to 120 days. Restricting paid sick leave to unionized workers only is a concern given that only 10% of the labor force is unionized [65]. Therefore, for most workers, regardless of unionization status, taking up maternity leave may imply forgoing income that may be critical for child nutrition and care. Just like sick leave, liability of paying maternity leave benefits is on the employer. In the past, liability was shared between the employer and the social security system. There are plans to introduce a maternity protection social insurance scheme which will, among other things, expand coverage of maternity protection and take up liability for paying benefits [66].

All employers are required to contribute to the Workers' Compensation Fund Control Board (WCFCB) to insure their workers against any injury or sickness that they may suffer as a result of their job. In so doing, the liability of paying for any disability is shifted from the employer to the social insurance scheme. The scheme compensates a worker for a disability suffered during the course of work for a maximum of 18 months, after which, if they do not recover, may be certified as permanently disabled and qualify for a lump sum payment or a pension for life depending on the extent disability.

These programs provide insurance against income loss. An important contributory social protection program that insures individuals against health expenditure risk is health insurance, which can either be voluntary or compulsory. Compulsory health insurance schemes are normally implemented through government legislation which enrolls and mandates all specified individuals to contribute to the scheme. On the contrary, enrollment and contribution to voluntary schemes–offered by private insurance companies or community based insurance schemes–are optional. Zambia does not have a compulsory or social health insurance scheme. Health insurance is offered by formal health insurance companies, were most of the insured individuals are from the formal sector. In most circumstances, employers normally take out group health insurance for their employees. Government efforts to introduce a national social health insurance scheme have been hampered by the size of the formal sectors, which in general is too small to generate sufficient funds for health financing. Notwithstanding this problem, plans to introduce a social health insurance scheme for the formal sector, and later extending it to the informal sector, have reached advanced stages.

The critical thing however is that the penetration of social insurance and labor programs, including NAPSA, WCFCB, maternity protection, and health insurance, is very low, with only 1% of the Zambian population being covered [48].

2.3.2 Non-Contributory Social Protection

Given that most of population is either in the informal sector or out of the labor force, social protection can realistically be extended to the majority through non-contributory schemes such as social assistance and empowerment programs. Social assistance refers to benefits that the government extends to sections of the population that meet specified levels of vulnerability. Such benefits include cash or in-kind transfers, fee waivers, subsidies, social spending on health and education, etc. Empowerment programs on the other hand involve providing access to services or work that would help individuals generate income, these include micro-credit, food for work initiatives etc.

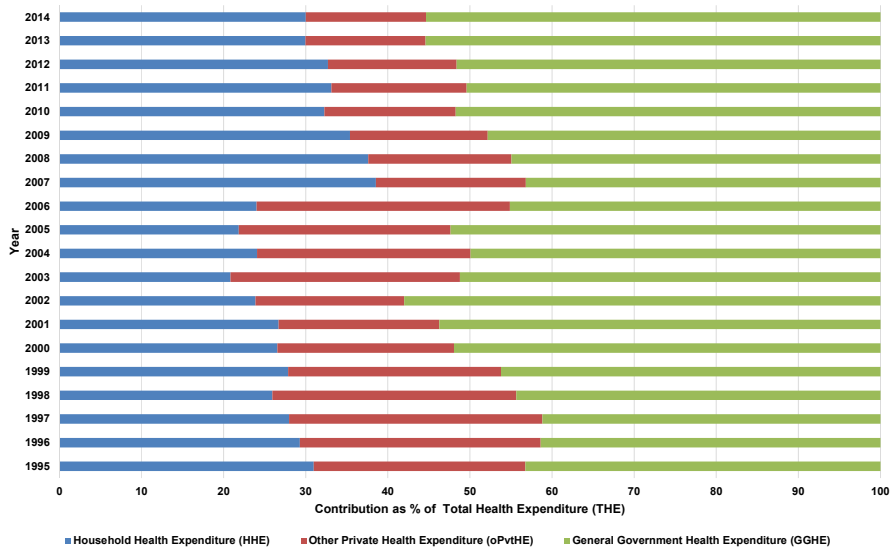
One of the most important fee waiver programs that the Zambian government implemented was the removal of user fees for health services. The goal was to reduce financial barriers in accessing health services, which would translate to increased utilization and, ultimately, improvement in population health. In 2006, user fees were removed in 54 rural districts and maintained in the other 18, designated as urban. The justification was that rural households

are more vulnerable because of higher poverty levels. There are suggestions that utilization of public health services may have increased following this wave of removal [67–69]. The user fee removal policy was extended to the rural areas of the 18 previously unaffected districts in 2007. No previous study has investigated the effect of this wave of removals. In 2013, user fees were removed in all public facilities throughout the country.

Despite the removal of user fees, household health expenditure does not appear to have responded as anticipated, even for individuals visiting public health facilities [70, 71], raising concerns of the extent to which the policy could have provided financial risk protection. Looking at how health care has been financed may give us a clue on how household contribution to total health expenditure faired before and after the removal policy.

Health care financing can be viewed as having three components; (1) general government health expenditure (GGHE), which is composed of government budget allocations to the health sectors and external (donor) financing (2) Households health expenditure (HHE), which is out of pocket spending by households excluding spending by private institutions and organizations, and (3) Other private health expenditure (oPvtHE), which is composed of spending from private organizations, e.g non-governmental organizations (NGOs) and employers, through local resource mobilization and external (donor) financing. National health accounts (NHA) data from the World Health Organization [50] shows that household health expenditure (HHE) as a share of total health expenditure (THE) actually rose in 2007, a year after the first wave of user fees removals. Although the share of HHE has averaged 29% over the whole period, 1995–2014, it soured to more than 30% beginning 2007 (Figure 2.2).

Figure 2.2: Health care financing in Zambia, 1995–2014**



**Authors own computation based on data from the World Health Organization [50]

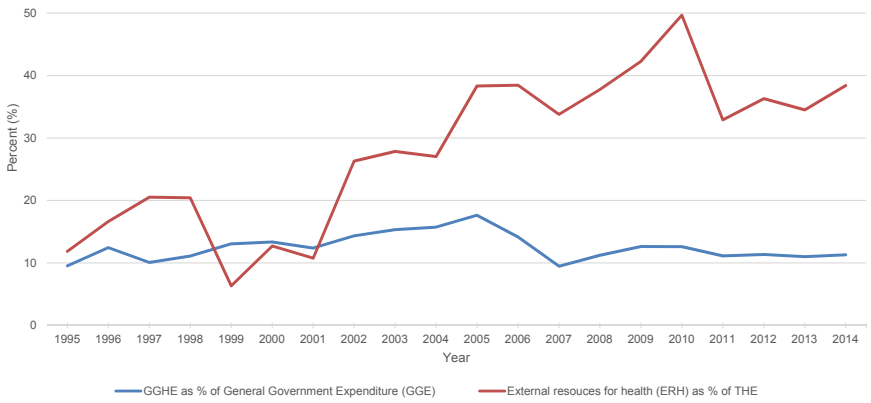
The high share of HHE can partly be attributed to the fact that the health system is poorly funded, which forces individuals, even those who utilize public facilities, to seek services and drugs from the private sector [72]. Inadequate funding of the health sector has resulted in heavy reliance on external financing. The bulk of these external funds go to government health budget support as official development assistance (ODA), forming part of general government health expenditure (GGHE). Figure 2.3 shows that external financing of health services rose substantially in the early 2000 from as low as 10% of THE to about 50% in 2010, before dropping to just under 35% in 2011.[3] This period of sustained increases in external financing coincides with the period of renewed global commitment to improve child health as well as combating infectious diseases such as HIV and Malaria under the millennium development goals (MDGs). It also coincides with the end of structural adjustment reforms (SAPs) and the beginning PRSP implementation. Zambia experienced substantial increases in child health interventions. Free HIV Treatment, through, Antiretroviral therapy (ART), was also introduced.

GGHE as a share of general government expenditure (GGE) responded favorably to the increase in external financing rising from around 12% in 2001 to over 18% in 2005. Improved external financing was associated with gains in key health indicators. According to the Demo-

[3]The drop was occasioned by the withdrawal of Swedish support due to a corruption scandal in the ministry of health [73]

graphic and Health survey [60], under five mortality rates dropped from an average of 168 per 1,000 live births in the period 1997-2002 to 75 per 1,000 live births in the period 2009–2013. Stunting for children under the age of 5 years also reduced from 53% to 40% between 2002 and 2014. There was also a reduction in adult mortality rate, from an average of 459 in the period 1994–2002 to 312 per 1,000 in the 2009–2013 period. HIV prevalence only dropped, though marginally, from 15.6% in 2001-2 to 13.3% in 2013. Notwithstanding the positive strides, these indicators still remain above the sub-Saharan Africa average. This is not surprising since increases in external financing was not matched by government funding. In fact, the share of general government health expenditure (GGHE) in general government expenditure (GGE) started dropping in 2006 and has since averaged 11%, falling short of the 15% target set during the Abuja declaration of 2001.

Figure 2.3: General governement expenditure on health and external resources for health, 1995–2014



****Authors own computation based on data from the World Health Organization [50]**

Apart from user fee waivers for health services and increasing health care spending owing to external financing, there are other non-contributory social protection schemes that have gained impetus as a result of both government and external support. The four main schemes, by coverage, are the public welfare assistance scheme (PWAS), the social cash transfer (SCT) scheme, the food security pack (FSP), and the school feeding program.

The public welfare assistance scheme (PWAS) is the oldest and largest social assistance scheme by coverage, operating in all 103 districts. Through in-kind support, PWAS aims

at providing assistance to the most vulnerable population in meeting their needs for health, education, shelter, and food. The scheme targets 2% of the Zambian population, according to it's definition of "the most vulnerable" although it reaches much smaller numbers [62]. A program that was initially nested in the PWAS but has evolved to become independent, and bigger, by expenditure size, is the social cash transfer (SCT) scheme.

As for 2016, the scheme provided a monthly unconditional cash transfers of sixty Zambian Kwacha (ZMW 60) to eligible households, an amount deemed sufficient to provide a monthly supply of basic food for very poor households. The pilot scheme was rolled out, in a staggered way, to 19 districts in the period 2003–2013 [74]. The objective of this pilot was to assess the effectiveness of different models of targeting the most vulnerable households. Each of the 19 districts implemented one of the following models: (1) The inclusive model (IM), which targeted the poorest 10% among incapacitated and destitute households (2) The child grant (CG), where households with children under the age of 5 years were targeted, (3) The multiple categorical (MC) scheme, which targeted households with a disabled member, and household headed by a female or elderly person with at least one orphan, and (4) The social pension scheme, which provided a pension to all people aged over 60 years in a district.

Evaluations of the child grant and multiple category schemes showed that cash transfers are effective in improving child health outcomes; reducing the incidence of sickness; improving nutrition and education outcomes; increasing maternal health care utilization; and strengthening household resilience to a number of shocks [75–77]. A further evaluation of the four targeting models recommended that the inclusive model (IM) was the most effective in identifying and selecting the poorest households [78]. The Inclusive Model (IM) was adopted, with a slight modification, as the method of targeting beneficiaries in the scale up of the SCT scheme following drastic increases in government funding. The SCT was scaled up to 31 more districts and the number of households expected to be covered increased from 61,000 in 2013 to 185,000 household in 2015 [79], which translates to 7.4% of households in Zambia. [4]

The food security pack (FSP) program is another important social assistance scheme. It provides vulnerable rural household with farming inputs so as to reduce food insecurity. Some of the eligibility criterion include having disabled members, orphans, females as household heads, orphans, and terminally ill-patients. Since terminaly ill-patients may require the care of household members who may need to take time off productive activities, the program provides

[4]The total number of households in Zambia is estimated at 2,500,000 according to the Living Conditions Monitoring Survey (LCMS) 2015 [51].

partial protection for health shocks of longer duration, which considering the high burden of HIV, is important. The challenge is that the coverage of the program is low. In reality, FSP is a prototype of the much larger farmer input support program (FISP). The aim of FISP is to reduce poverty in general by improving the availability of subsidized inputs to viable small scale farmers. However, evaluations have shown that in most cases, the benefits of FISP accrue to the well off , and yet it gobbles as much as 73% of the agriculture poverty reduction budget[80]. On the contrary, FSP, which is said to target the poor better, is only 5% of the FISP budget. There are concerns that too much spending on a program that benefits the well off explains the persistent levels of poverty and vulnerability in rural areas.

Education is another important avenue of improving population health, reducing vulnerability, and breaking the intergenerational transmission of poverty. There are two important social assistance schemes that have been implemented in order to directly improve education, particularly, school attendance. First is the removal of school fees for basic education (First to ninth grade). Although this policy has been associated with remarkable increases in enrollment, measures of performance show that quality of education offered needs improvement [81]. It is important to note that only quality education can meaningfully improve cognitive and non-cognitive skills, which are critical to health production.

The second program designed to improve school attendance is the home grown school feeding (HGSF) program which aims at minimizing hunger among school children. It is recognized that hunger may affect mental alertness, affect health, and lead to frequent absence from school as children work or help their families in looking for food. Implemented as a joint effort between the World Food Programme (WFP) and the Ministry of Education from 2003, the scheme provides a meal to more than 800,000 children throughout the country each day. It also provides a local market to farmers products [82].

In summary, social assistance programs have received unprecedented attention over the past few years in Zambia. Social assistance can improve a households resilience to health shocks. For example, households who received a cash transfer in Zambia reported using their saving from the scheme to protect themselves against shocks [76]. However, although more than 63% of the population live in poverty, only the poorest of the poor are targeted by various social assistance schemes because of budgetary constraint. As a result, it is unclear how resilient the large majority of household are to shocks of different kinds. In Zambia, the most common and serious shock to a household is a health shocks [76, 83]. Given the limited

coverage of social protection schemes, questions remain on what coping strategies household employ when hit by shocks.

Chapter 3

Objectives, Rationale, and Conceptual Model

3.1 Objectives and Rationale

3.1.1 Overall Objective

To examine the extent to which households are protected from the welfare effects of health shocks, and evaluate some policies meant to minimize vulnerability in the general population and in specific groups such as children and individuals from low socioeconomic background.

3.1.2 Specific Objectives

This thesis consists of the three objectives, each of which is a sub-study translating into a paper. The three objectives are:

1. Assess the effect of health shocks on household consumption, income, and health expenditure, as well as the extent to which households use borrowing and selling assets as self insurance strategies in the absence of complete social protection systems, during and after structural adjustment reforms (SAPs).

2. Evaluate the short and long term effects of the user fee removal policy on health expenditure and overall utilization of health services, and determine heterogeneity in the utilization response.

3. Investigate determinants, and changes in inequality, of stunting and fever among children below the age of 5 years over the period 2007–2014.

3.1.3 Rationale

An investigation of the effect of health shocks, as we do in the first sub-study, is important in determining the desirability and appropriate design of social protection policy [84, 85]. By examining heterogeneities, it is possible to determine which sections of the population may require increased social protection. To be most useful, such an investigation needs to consider how health shocks simultaneously affect multiple outcomes (consumption, health expenditure, labor income, and self insurance strategies). The reasoning behind this assertions is worth explaining. Since household economic welfare is determined by consumption of goods and services, welfare losses may result if health shocks cause significant variations in consumption. In this case, resources spent to protect households from the effect of health shocks would improve household welfare. On the other hand, if households are able to protect themselves from such shocks, social protection systems would be inefficient as they crowd out self insurance mechanisms. If however households protect their consumption using inefficient strategies that may affect their welfare in the long term, social protection may still be desirable even if consumption does not move, or moves little. Reaching this conclusion may however require more information. Since, health shocks affect consumption through increased medical spending and reduced labor income, there may be need to assess if these variables respond much more than consumption does in order to conclude that there may be self insurance, which further raises questions on what type of self insurance mechanisms the household could be employing. For example, once one finds that health shocks substantially increase health expenditure (medical spending) and reduce earned income but that consumption does not move or moves far less, suggesting self insurance, one may go on to assess the type of self insurance strategies used in order to determine if protecting consumption is achieved at the expense of household long term welfare.

Despite the importance of assessing the effect of health shocks on all these outcomes (consumption, health expenditure, labor income, and self insurance strategies) simultaneously, we are not aware of any study that has undertaken this task in sub-Saharan Africa. Previous studies have focused on the effect of health shocks on one or two these outcomes and the findings are mixed [29, 86–92]. The other challenge with most of these studies is that they mostly focus on rural areas. It is not clear whether such findings would hold in urban areas or in the general population. There is also a gap in understanding the effect of morbidity measures of health shocks such as sickness and injury since most of the studies in sub-Saharan Africa

have focused on one health shock measure, namely, death. Studies in Asia encouragingly examine a number of health shocks, including morbidity shocks, and in the general population [5, 7, 8, 93–96]. However, our discussion in Section 1.2 makes it clear that Asia and sub-Saharan Africa may face different circumstances which makes it difficult to assume that findings from Asia would carry over to sub-Saharan Africa.

Our study is the first to examine the effect of health shocks on all four outcomes simultaneously, in the general population and using a measure of morbidity, injury. Looking at the general population also allows us to control for district effects in our model, effectively questioning whether social protection or risk sharing is well developed at the district level. Most studies test the hypothesis of consumption insurance at the village level, e.g [8, 95]. However, in our case, and indeed in a number of sub-Saharan Africa countries, social protection schemes, health services and other programs are delivered at the district level and not the village. Moreover, districts are quiet homogeneous and it is thus possible to speak of risk sharing networks within districts.

The second sub-study evaluates an important social protection policy which removed user fees for health services. This was done with an objective of protecting households in two dimensions. First, the removal of user fees would increase utilization of health services which may ultimately reduce health risk or improve health. Second, the removal policy would reduce health expenditure, and contribute to providing financial risk protection (reducing household health expenditure risk). Although a number of studies have assessed the effect of the removal of user fees, almost all these studies focused on changes in utilization of public health facilities, and not in overall utilization of health services [33, 67, 97]. It is important to note that changes in utilization of public health facilities may come as a result of switching from private to public facilities which may imply that the policy only succeeded in crowding out private health services. Given the financing challenges of government health services, such an outcome would imply that resources would be better spent elsewhere. Moreover, if the changes in utilization where due to switching, health improvements may not materialize as envisaged by the policy.

Additionally, access to household data enables us, unlike previous studies, to examine how the removal policy affected utilization of individual from different socioeconomic backgrounds. This is important because the policy was meant to elicit a greater utilization response from individuals who are likely to face substantial challenges in accessing health services.

Previous studies have also not looked at the effect on household health expenditure, which is clearly an outcome of policy interest. Exceptions to this are cross sectional studies from Uganda [72, 97]. Our study is important because it uses a natural experiment to evaluate these effects. Moreover, given the staggered nature of the removal policy, we are also able to determine whether or not the observed effects were sustained in the long term. Previous studies have suggested that effects of the removal policy are not sustained in the long term. However, since most of these studies faced data or methodological limitations, the long term effect of the removal policy remains uncertain.

Based on the premise that high health risk may negatively affect household welfare, reducing it is an important social protection goal. More important however is reducing health risk in the most vulnerable populations, especially those who have no control on circumstances that would improve or damage their health. One section of such a population is children. In short, it is important to reduce inequality in childhood ill health. In order to reduce inequalities, it is cardinal to examine not only factors that determine child health but also the factors that may contribute to the changes in inequality at any given time. This is very important if policy has to curb increases in inequality.

While it is known that the substantial increases in funding and scale up of child health interventions, in the period leading up to the 2015 target for the Millennium Development Goals (MDGs) was associated with improvements in child health, and rising inequalities, we are not aware of any study that has examined the factors that could have been driving the observed increases in childhood health inequalities. To undertake this analysis, we use two important measures of childhood ill health, stunting and fever. Stunting is a broad measure of child under-nutrition and may also represent an accumulation of health shocks both in-utero and in early childhood. It is reported that nearly half of under-five deaths are related to under-nutrition [45]. Fever on the other hand may encompass a number of illnesses such as malaria, as well as bacterial and viral sickness. Our study may be important in informing policy markers on what factors to focus on as they brace themselves to tackle inequality under the post 2015 development agenda, the Sustainable Development goals (SDGs).

3.2 Conceptual Models

Several conceptual models may be useful in understanding mechanisms behind the findings of this thesis. The first model, human capital theory, relates to all the papers (Paper I, I, and III) while the second one, theory of full consumption insurance, mainly relates to Paper I. Because these are standard models, we do not show their formal setup and the testable implication.[1] We state however that our empirical models can be shown to be reduced form specifications from these models.

3.2.1 Human Capital Theory

Human capital theory is the main conceptual framework of this thesis. The basic idea in human capital theory is that an individual can be viewed as having a specific stock of human capital or capabilities, such as health, knowledge (education) and wealth, which, like any other capital item, produces a flow of services (benefits) enabling one to be productive in the market sector where they produce money earnings or in the household where they produce other commodities, e.g. child rearing, that feed directly into their utility function. An individual therefore has an incentive to invest in human capital by, for example, utilizing health services or schooling. The cost of investment is direct money outlays and the opportunity cost of time used in investment instead of earning. A decision maker weighs the benefits and costs of investment in determining the optimal holding of the capital stock.

We first discuss the general life cycle model of health capital, or more precisely the adult life cycle model of health capital. It is most useful for understanding health investment in paper I (changes in medical expenditure following a shock) and in paper II (changes in both utilization and medical spending induced by a perceived reduction in health investment cost). It also helps us to appreciate how individuals (and policy markers) care about health stock, for it's on sake, and attempt to minimize shocks to it (Paper II and III) because shocks may affect earnings (Paper I), or increase health investment costs, e.g health expenditure (Paper I and II), both of which would make the lifetime budget constraint more binding, and lower their utility or welfare through reduced consumption (Paper I). This motivates investments in health (Paper I, II, and II).

[1] A formal discussion of the theory of human capital can be found in Becker [98] and Grossman [99]. Heckman [100] discusses the childhood human capital model of capability formation and Cochrane [101] discusses the theory of full consumption insurance.

Although lower childhood health stock, and health shocks, may affect labor earnings of adult care givers, the motivation in child health investment may not be for an immediate earnings effect, but for childhood later life outcomes. We thus present a separate model of childhood health investment or production to appreciate how childhood health and inequalities may be produced. This is also important in selecting appropriate variables for the model and understanding unobservable factors that may reinforce inequalities. Health inequalities may persist because different observables (e.g health) and unobservables capabilities (e.g cognitive skills) reinforce each other and are produced differently. Thus, for childhood, we consider a general theory of human capital/capability production, which captures both health and other capabilities, in children and their caregivers.

Life-Cycle Model of Adult Health Capital

Human capital theory has received so much interest since the pioneering work of Becker [98], Ben-Porath [102], Mincer [103] and Becker et al. [104]. While it was clear from these studies that health has an effect on earned income, it was Grossman who explicitly modeled the demand for health capital [99, 105]. Since this work, important extensions have been made [17, 106–108].

In modeling the demand for health, it is assumed that an individual derives utility from consuming health and market goods. They solve an optimal demand or consumption path for health and other market goods that maximizes their welfare subject to a lifetime budget constraint. The consumption of market goods, through the budget constraint, partly depends on earnings, which in turn is determined by, among other things, the optimal stock of health, and shocks to it. Thus, Health is consumed and is also used to produce earnings, implying a consumption and investment benefit of health. Consumption benefit because it enables the consumer to enjoy 'pain' free days and an investment benefit because less sick days gives the consumer more labor time to work in the labor market and earn more. Apart from increasing labor time, having a higher health stock also enables the consumer to be more productive and hence increase his earnings further. Earnings, together with exogenous wealth, constitutes life time income. Thus, health is clearly predicted to affect lifetime income–which through the lifetime budget constraint also affects the consumption of market goods and services. A reverse causality from income to health is also possible. Individuals with lower lifetime budget constraints may be better able to buy market commodities for health investment. Alternatively,

given their income levels, they can afford time to invest in health behaviors [17, 108]. In view of the benefits of health and its possible effect on welfare/utility (consumption), a consumer desires a higher health stock and invests in it.

The overall level of health stock at any given time period is determined by the previous period's health stock, health investment or repair, and depreciation of the health stock. The first determinant of health implies that health production is a dynamic process so that lower health stock (or health status) in the current period may lead to poorer health in the next period. In this case, individuals with chronic conditions may be more likely to face health shocks which lowers their stock further, or even drop it to zero (death).

In terms of health investment or repair, individuals undertake investments (one of which is healthcare) in order to raise their health stock or repair it following a health shock. Health shocks may temporarily or permanently lower the stock of health [106]. Generally, investments in health may require both time and market inputs, e.g time to go to a health facility plus payment for health services, time to exercise plus payment to the gym, etc. In a simple model, health services are assumed to be the only inputs in the health investment function. In reality, there are many other inputs. Regardless of the current levels of investment, a health shock will increase utilization of health services and the associated spending. In the same vein, lowering the cost of health services and the time required to get these services (e.g waiting times)–i.e. reducing the cost of health repair–increases utilization of health services, which, according to the model, increases health stock.

Other capabilities or forms of human capital such as education, cognitive and non-cognitive abilities (e.g time preference) increase the efficiency of health investments. The model predicts that Individuals with higher education, lower time preference (who are patient), or with more perseverance will invest more in health and in a more appropriate manner, e.g, they will follow a doctors prescription, not give up on medication, etc.

While health investments add to the stock of health, depreciation reduces it. Depreciation of the health stock can be gradual or sudden, in which case a health shock is said to have occurred. Specifically, health (the body) depreciates as a function of age as well as on how it is used in (1) consumption, e.g. in consuming unhealthy foods, behaviors, etc, and (2) work, e.g. physically unhealthy working conditions. All things equal, age is associated with lower health stock levels. Individuals involved in occupations that strain the body or that are physically damaging will have higher levels of depreciation [17, 108]. For example, they may be more

prone to injuries or occupational diseases such as chest infections.[2] Individuals with unhealthy consumption and behaviors such as smoking, drinking or consuming rich foods will have lower levels of health, all things equal [109]. This also implies that apart from health affecting consumption through it's effect on productivity and earnings, consumption may affect health. [3]

Life-Cycle Model of Child Health

The key points in the preceding discussion also hold in childhood, but with two key differences. First, most of the investment in childhood health is not done by children themselves, but their parents or caregivers. Hence, differences in parental investments and circumstances largely determine inequalities in childhood health, which persist later in life. Second, as mentioned earlier, as opposed to investment in adult health which may be partly motived by both short and long term earning effects, investments in childhood health is not motivated by a short term earning effects, but on later life outcomes. To fix ideas, we informally discuss a simple model of childhood capability (human capital) production [100].

Children posses a vector of capabilities, which include health stock, cognitive abilities, and non-cognitive abilities, e.g time preference, self control, risk taking, etc. These capabilities are produced by a child's environment, her capabilities in the previous period, and her parents' capabilities. Inequalities in childhood capabilities in the population are due to differences in the level and effectiveness of parents' capabilities as were as differences in environmental exposure. Thus, some parents have disproportionately higher levels of capabilities which result in inequalities in child capabilities. Not only are the parental capabilities unequally concentrated in the population, but their effectiveness in producing child capabilities are also different. For example, while a parent with a larger wealth capability may be able to produce higher child health capabilities, it may also be the case that the effect, in terms of elasticity, of this parent's wealth capability on her child's health may also be higher than that of a comparable parent.

[2] Case and Deaton [17] examined the effect of occupation on health. They found that a substantial amount of life cycle differences in health between individuals in the top income quartile and those in the bottom quartile were a result of low income individuals undertake physically taxing occupations. They further found that the effect of income on health is dramatically reduced when the role of occupation is accounted for. It is thus crucial to control for occupation in our empirical models

[3] van Kippersluis and Galama [109] modeled unhealthy consumption as a determinant of depreciation of the health stock and showed that social-economic differences in health between wealthy and poor individuals may be explained by the fact that wealthy individuals have lower unhealthy consumption as the health cost of unhealthy consumption increases with wealth. Higher education and other cognitive as well as non-cognitive abilities are also associated with lower health depreciation.

The differences in the effect of wealth capabilities may be due to other unobservable parental capabilities such as cognitive skills, time preference, self control etc. Thus, differences in levels and effects of parental capabilities explain differences in child capabilities. There is evidence of intergenerational transfer of capabilities from parents to children [110, 111].

An important aspect of this model is that inequalities in child capabilities may persist because children whose parents have more or higher capabilities gain a number of capabilities at the right time—mostly in the early stages of their life cycle–and these capabilities reinforce each other at any stage, and across different stages of the life cycle. Specifically, production technology is assumed to be different throughout the life cycle of a child. Some capabilities are more effectively and efficiently produced at certain stages of the child's life cycle. For example, health,measured as adequate height for age, may be more efficiently produced in the first two years of life. Similarly, the productivity of inputs from parental capabilities (e.g income, education, genes, etc,) and the environment (e.g community child care policies, breastfeeding promotion) differ over the child's life cycle, implying that it may require more inputs to produce the same capabilities if critical stages are missed. Beginning at birth when a child receives an initial endowment of capabilities,[4] child capabilities formed at any stage earlier in the life cycle play a role in producing other capabilities at that stage or in later stages through cross productivity, self-productivity and dynamic complementarity effects.[5] These effects reinforce inequalities produced by parental capabilities and the environments.

3.2.2 Theory of Full Consumption Insurance

Human capital theory clearly predicts that health shocks will affect earned income, medical spending and ultimately consumption, by making the lifetime income constraint more binding. On the other hand, the theory of full consumption insurance holds that idiosyncratic risk, like health risk, should not affect household consumption when it materializes because households dislike risk and seek ways of insuring against it, whether through social protection systems or

[4] Child's initial endowment of capabilities is determined by the production technology in-utero which is governed by genetic expressions, parental investment during pregnancy (e.g diet, health behaviors, smoking etc) and in-utero exposure to environmental insults.

[5] Cross-productivity effects imply that different child capabilities in any given period affect their ability to acquire other capabilities in that period. For example children in better health may acquire cognitive skills more easily. Self productivities imply that childhood capabilities developed in an earlier period (e.g non cognitive skill such as risk taking) have a tendencies to reinforce capabilities in the next period (e.g health). Dynamic complementaries suggest that capabilities produced in an earlier period increases the productivity of investments in subsequent periods so that fully reaping the benefits of investment of child capabilities may require continuous investment.

self insurance strategies.

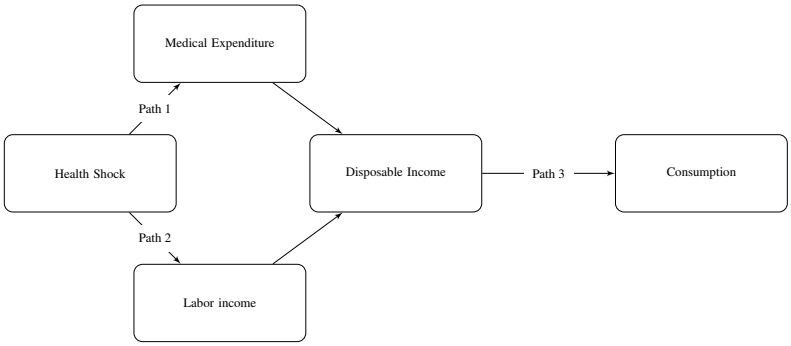
Specifically, the full consumption insurance theory says that individual household consumption should not be responsive to idiosyncratic shocks (such as health related income shocks) if, for any group of households, an arrangement exists that achieves a full Pareto optimal allocation of risk [95, 101]. Such an arrangements may include: (1) complete markets where individuals can trade all possible risks; (2) institutions such as insurance companies, banks and other financial institutions (3) central planners such as governments ; and (3) Informal risk sharing arrangements (for example at community level) or households' own self insurance or coping strategies. Such arrangements, for any given number of households, provides protection to households in a novel way: it is as if households surrendered all their resources to a central planner who later redistributes them to households irrespective of their current resources (which may have been affected by health shocks). This makes consumption non-responsive to individual household health related income shocks [101]. In the absence or inadequacy of these risk bearing arrangements, health shocks may affect consumption.

Thus, a fully functioning social protection system can be viewed as an insurance system where individuals contribute to the central pool through taxes, levies, premium, e.t.c, and the planner uses these resources to reduce population health risk and insure against its effects. Reducing health risk would involve investing some of these resources in public health prevention or curative interventions so that individuals suffer less health shocks and thus, avoid their welfare effects. Insuring against the effects of health risk may involve the central planner or government transferring resources that individuals contribute only to individuals affected by health shocks so as to keep unchanged their welfare or consumption.

Clearly, our discussion of the insufficiency of the social protection system in Zambia makes the possibility of, even modest, consumption insurance unlikely. Thus, we anticipate that, since households dislike risk, they will adopt some self insurance strategies. If households are constrained in the use of such strategies however, their consumption will ultimately be affected, basically through two pathways, increased medical spending and reduced earned income. Figure 3.1 shows that health shocks may affect consumption through their effect on disposable income. Disposable income could be reduced via path 1 or path 2, or both. Path 1 is a reduction due to increased medical expenditure while path 2 is due to reduced labor income.[6]

[6]There is also a possible direct effect of health shocks on consumption, known as state dependence, which we do not show in this figure.

Figure 3.1: Transmission mechanism of health shocks



However, whether or not consumption will be affected depends on the existence of formal or, in there absence or inadequacy, informal risk bearing arrangements at each of the three paths in the figure. Formal risk bearing arrangements include health insurance, disability insurance or sick leave [5], financial institutions providing credit , social welfare programs such as cash transfer. On the other hand, informal risk bearing arrangements community risk sharing arrangements and self insurance strategies such as selling assets, informal borrowing, pulling children out of school, shunning medical care, and within household labor substitution.

Chapter 4

Methods and Results

4.1 Methods

4.1.1 Overview

This thesis comprises three sub studies, or papers, corresponding to the three specific objectives in Section 3.1.2. Methods used for each paper are summarized at the end of this chapter in Table 4.1. More detailed information is in the respective papers. We provide brief descriptions below.

4.1.2 Methods: Paper I

Data

We used data from the 1996, 1998, 2004 and 2006 waves of the living conditions monitoring survey (LCMS) to assess the effect of health shocks on consumption, income, household health expenditure, and coping strategies. The LCMS is the single largest and richest source of nationally representative household data capturing both measures of health and household living standards such as consumption, income, and production, among others. It uses a two stage cluster sampling design. In the first stage, geographical clusters, called Standard Enumeration Areas (SEAs), are drawn. Each SEA is then stratified according to economic activity or living standards to ensure representativeness. Households are then sampled using a systematic sampling method after which a questionnaire is administered by trained enumerators. Our health information of interest was collected by asking whether any member of the household was sick or injured in the two weeks prior to the survey. Only when a household reported any

injury do we consider that household to have experienced a health shock.

In our data, food and non-food consumption have a reference period of one month. We computed both totals and shares of food and non-food consumption. Health expenditure also has a one month reference period and is not included in our measure of non-food consumption. Similarly, all income categories, with the exception of farm income, have a one month reference period. We used this information to define two measures of income, unearned income (sum of remittances received, grants received, and borrowing.) and non-farm earned income (sum of wage and non-farm business income). We excluded farm incomes because they had a one year reference period, which may reflect production way before the health shock occurred. To account for inflation, all income and expenditure data were adjusted to the common base year of 2006 by using the Consumer Price Index (CPI).

Lastly, we consider data on household self insurance strategies. In the surveys, households were given a list of possible coping strategies and then asked whether or not they used any of the listed strategies when faced with hardship. The question did not relate specifically to health related hardships, but, like Sparrow et al. [7], we assumed that coping strategies for hardships in general are also relevant for health related hardships. However, the 2004 data for coping strategies was not available. Hence, we pooled all the surveys together for coping strategies because using the 2006 data only for the period after 2002 would have not allowed us to exploit cohort variation.

Empirical Strategy

Given that our data is from repeated cross-sectional surveys, our empirical strategy relied on minimizing possible bias that may arise from the fact that individuals who experienced health shocks may have been different in many unobservable ways to those who did not. First, we used injury as our measure of health shock. Injury, unlike sickness or death has the advantage of being unanticipated and accidental [112]. Sickness or death, as we saw from the human capital model in Section 3.2.1, may be more likely for individuals with low health status, and whose outcomes, e.g incomes, may have already adjusted, making it difficult to identify any change following a shock [90].

Though unanticipated and accidental, injury is unlikely to be exogenous nonetheless. The likelihood of injury may differ systematically by socioeconomic status, (e.g occupation,wealth, education etc), demographics, and residence. We controlled for these and other factors in the

regression model. Additionally, individuals with certain unobserved fixed effects such as risk taking attitude may also be more likely to be injured. We control for smoking and alcohol consumption given that these factors may modify risk taking altitude and injury is likely to be very correlated with alcohol and smoking behavior. The remaining fixed effects can be differenced out if one has access to panel data. Since we don't have panel data, we took advantage of the availability of repeated cross sectional data to minimize possible bias from the remaining fixed effects. The strategy was to use head of household cohort effects as a proxies for household fixed effects. [1] We also included fixed effects for the 72 districts. This is important in accounting for the fact that injury may be more likely in certain districts than others, e.g district with poor road infrastructure and where local safety laws are not strictly enforced. In addition to being the center of local government administration, and distribution of most services, social protection systems, especially social assistance programs are organized around districts. Districts are also relatively homogeneous and households may come together to form risk sharing groups. We can therefore talk of risk sharing within districts and postulate that if social protection systems are well developed, health shocks should have no effect once we control for district effects. We also controlled for district by year effects to account for time varying district effects.

In the estimation, consumption and income was each broken into two components: food and non-food consumption; and earned and unearned income. This is important because health shocks may affect components of consumption or income differently. For example, a health shock may reduce daily business income (earned) but increase unearned income if the affected household is given some money to help in recovery by friends or family [5]. For consumption, there may be substitution between food and non-food items depending on what consumption items are of greatest urgency in promoting recovery. If consumption is combined, these effects may cancel each other. Thus, our empirical model consisted of three systems of equations; namely, consumption (food and non-food), income (earned and unearned), and self insurance (coping) strategy systems. It is however important to consider the fact that the same unobservable factors, such as tastes and preferences, may be influencing food consumption and non-food consumption. Similarly, unobservable factors that may influence one to sell their

[1] The reasoning is based on the literature on pseudo panel regression models. In this framework, as long as one has access to repeated cross sectional data, a panel data can be formed by grouping all households based on the birth cohort (year of birth) of the head of household and collapsing all the variables into averages of these cohorts. Importantly, the cohort fixed effects averages the unobservable fixed effects of the heads of household, and provided there are no composition changes overtime and the surveys are nationally representative, the cohort fixed effect are consistent estimators of household fixed effects

assets when hit by a shock may also influence their borrowing behavior. Thus, we simultaneously estimated each of the three systems of equation, that is, consumption, income, and coping strategies, in a Seemingly Unrelated Regression (SUREG) framework. This allows for interdependency in the unobservable factors of each system, and leads to efficiency gains [113].

The effect on health expenditure was estimated using generalized linear models (GLM) with the appropriate family and link chosen using the modified park test and box cox test, respectively. These tests led us to choose a GLM Poisson model with log link. GLM has been found to be a more appropriate way of modeling outcomes, such as medical spending, that have a skewed distribution.

The problem with the above analysis is that we are controlling for household head cohort effects and other household head characteristics while the health shock measure is for any injury in the household, and not the household head. Thus, it may turn out that individual specific observable characteristics were actually the ones driving the outcomes. We ran an alternative model restricting injuries to the head of household to assess this possibility.

The analysis was conducted separately for the period before and after 2002. This is an interesting comparison because the year 2002, as we have thoroughly discussed in Section 2.2, marks an end to a period of numerous structural adjustment reforms (SAPs) spanning economic, social and health sectors. For each sub-period, we estimated the models in the full sample as well as in quartiles of per capita household consumption. In all models, we accounted for survey design in order to get the correct estimates and standard errors. The full set of controls in the regression models included year of birth cohort effects, district and district by year effects, total household consumption, household size, dummy for any consumption of alcohol, dummy for any smoking, residence (urban/rural) and head of household characteristics such as occupation, age, educational level, and marital status.

4.1.3 Methods: Paper II

Data

To evaluate the effect of the user fee removal policy in public facilities, we used data from five waves (1998, 2002, 2004, 2006 and 2010 waves) of the LCMS. Prior surveys could not be used because the number and boundaries of the districts were changed making it difficult to isolated treated from untreated areas.

Our interest is on health service utilization and expenditure for individuals who reported being sick or injured in the two weeks prior to the survey. Individuals who reported being sick or injured in the two weeks prior to the survey were asked if they consulted any health facility, and the type of health facility they consulted. From this information, we constructed three measures of utilization and two measures of health expenditure.

Our main utilization measure was used to assess the change in overall use of health services and was coded 1 if an individual reported consulting any health facility (private or public) and zero if not. The other two measures were meant to assess switching between private and public facilities. Specifically, our second measure took the value of 1 if an individual reported consulting a public health facility and zero if not. The third measure of utilization was defined to take the value of 1 if an individual consulted a private health facility and zero otherwise.

Health expenditure was defined at the extensive margin-whether or not somebody incurred any spending–and the intensive margin–the amount of spending incurred given that they had positive spending. Thus, our first measure of health expenditure was a dummy equal to 1 if an individual incurred any spending while the second measure was conditional health expenditure.

Other variables of interest included distance to the nearest health facility; household size; age and sex of the individual; and head of household age, sex, educational level, occupation, and marital status.

Empirical Strategy

We defined the first wave of user fee removal in the 54 out of 72 districts in early 2006 as treatment one (T1) and the individuals who lived in these districts as the T1 group. This enabled us to identify the short term effect of T1 in late 2006 and it's long term effect, in 2010. The second wave of the removal policy which extended the removal of user fees in 2007 to the rural areas of 18 districts that were previously unaffected was defined as treatment two (T2), and the individuals who lived in these areas were designated as the T2 group. This wave of removals enabled us to identify the effect of T2 in 2010. Since only urban areas of the 18 districts were unaffected by the removal policy, these areas formed our control group.

We applied difference-in-difference models with standard errors clustered at the district level to identify the effect of the user fee removal policy. The models included fixed effects for the 9 provinces to account for unobserved regional fixed effects that may have affected

utilization and health spending. We also included province by year effects which capture time varying unobserved regional effects. A full set of year effects was also included in the model as well as other control variables as discussed in Section 4.1.3. The base year was 2004. Our identification strategy relied on the common trends assumption which says that in the absence of the removal policy, outcomes in the treated groups would have followed the same trend as those in the control group, so that departure from trend is counted as the effect of the removal policy. This assumption is more plausible if outcomes in treated and control groups followed similar trends before the intervention. In our case, all pre-treatment outcomes in treated and control groups appeared to follow similar trends. We also formally tested this assumption by conducting placebo tests were pre-treatment interaction effects were included in our models and checking that they were zero. We also carried out a number of robustness checks to ensure that our results were not driven by other things happening other than the removal of user fees.

In the empirical set-up, the short and long term effects of the removal policy on all three measures of utilization (overall, public, and private utilization) was assessed using linear difference-in-difference models. We also investigated how overall and private health facility utilization of individuals from different socioeconomic backgrounds were affected by the first wave of the removal policy (T1). To do this, we ran linear difference-in-difference models on the categories or sub-samples of each of our two measures of socioeconomic status, occupation and education of the household head.

The empirical set-up for assessing the effect on health expenditure is not straight forward because of the large number of zero expenditure and particularly right skewed distribution. While one can use a logarithmic transformation and run ordinary least squares (OLS) regression on transformed data, this typically results in inefficient estimates due to the problem of heteroskadasticity. Moreover, this transformation excludes zeros from the analysis which could further introduce bias. Generalized linear models have been found to perform well in dealing with skewness in health expenditure data. To deal with both zeros and skewness, we combined the GLM and Probit models using the two part model (TPM). In the TPM, the first part looks at the effect of the policy at the extensive margin using a probit difference-in-difference. Due to the non-linearity of probit difference-in-difference, we computed marginal effects appropriately as discussed in Frondel and Vance [114]. In the second part, the effect of the policy on the intensive margin was assessed using GLM. In the GLM framework, one has to chose the link function and family that best suits the data. We used the Box Cox test

to chose the link function and the modified Park test to select the appropriate family. This gave us the gamma family with logarithmic link. The effect on unconditional spending was assessed by combining the extensive and intensive margins of the TPM. All standard errors were estimated using a bootstrap procedure with 1,000 replications.

In all our models, standard errors were clustered at the district level.

4.1.4 Methods: Paper III

Data

To study childhood stunting and fever during the period of massive scale up of child health interventions, we used data from the 2007 and 2014 waves of the demographic and health survey (DHS). The DHS is the richest and largest household survey on health in Zambia, and many other low income countries. In Zambia, it is conducted by the central statistical office (CSO), MEASURE, and other collaborating institutions.

The DHS uses a stratified two stage cluster sampling design. Each province (9 in 2007 and 10 in 2014) is divided into two strata; urban and rural. From each stratum, SEAs are drawn with probability proportional to the size of the SEA, which is the number of households in that SEA. The second stage selects households using a systematic sampling method with all household having the same chance of being selected. After selection, three types of questionnaires, namely, household's, woman's, and man's questionnaire, are administered to eligible members of the household by trained enumerators. We were interested in information on children under the age of 5 years, which is captured in the household and woman's questionnaires. From this information, we constructed two dummy variables, fever and stunting, which are our outcomes of interest.

To collect information on fever, mothers were asked if their child had fever in the two weeks prior to the survey, to which they had to respond yes or no. On the other hand, stunting information was based on measured heights (by trained interviewers) of all children under the age of five years in eligible households. We classified a child as stunted if their height-for-age z-score (haz) was less than two standard deviations of the reference population based on the WHO 2006 growth standards.

Other variables used in the analysis included child's birth weight, age in months, birth order, place of delivery (facility or home) and duration of breastfeeding; mother's height, weight, educational level, age, and employment status; and household characteristics such as

residence (rural/urban), access to improved water source [2], whether or not a household had an improved toilet, household size, number of children under five years, and a wealth index. The wealth index is computed using principle component analysis (the first principle component) and is provided together with the data.

Empirical Strategy

We used random intercept (multilevel) linear models to examine the determinants of childhood stunting and fever, and whether or not the neighborhood were a child lives is important in explaining childhood ill-health. Thee models had two levels, the first of which was at the child (individual level) while the second was the Standard Enumeration Area (SEA). We did not add any covariates at the second level. Our interest was in computing the intra-cluster correlation which helped us to assess whether unobservable factors that affect childhood ill-health at the SEA level were as important as individual or household level factors. If SEA level variation of unobservables is significant, one can conclude that there was spatial clustering in childhood ill-health, a form of geographical inequality.

To quantify the extent of socioeconomic inequality in stunting and fever, we used the concentration index (CI). In our case, the CI measures the extent to which childhood stunting and fever depends on the wealth of a household. It stretches from -1 to 1 and it is zero if there is no dependence, and hence no inequality. If all the childhood ill-health is concentrated on children from households with lower wealth, then the CI will be -1. It will be 1 if all the ill-health is concentrated on the well off.

However, the standard CI understates the extents of inequality (the bounds may not be -1 and 1) if the outcome measure is binary, as it is in our case. Thus we used a correction method proposed by Wagstaff [115]. For each measure of childhood ill-health, we computed the corrected CI for 2007 and 2014. Next, to compute the change in inequality between the two periods for each measure of childhood ill-health, we subtracted the 2007 CI from the 2014 CI.[3] The computation of the corrected CI and the change over the period involves multiple steps, each of which introduces uncertainties. Hence, analytical standard errors from the last stage would make confidence intervals appear narrower than they actually are. To guard against this, standard errors were computed using a bootstrap procedure with 1,000 replications.

Having computed the change in CI of each childhood ill health measure, we wanted to un-

[2] The definition of "improved" is consistent with what is used in the DHS [60].

[3] We refer to the corrected CI simply as concentration index (CI)

derstand which determinants had the greatest contribution to the change in the CI. To do this, we decomposed the change in the CI between 2007 and 2014 for each measure of childhood ill-health using the Oaxaca type decomposition [116]. In this approach, the contribution of each determinant is due to two main sources. First is the change in the CI of the determinant. If inequality of a determinant over the period 2007–2014 changes, as measured by its CI, inequality of childhood ill-health will also change. Second is the change in the effect of the determinant, measured as an elasticity. Whether or not the change in the effect of a determinant will increase or reduce inequality dependents on how unequally distributed the determinant is in the first place and whether or not the determinant is protective (associated with good childhood health). If the determinant is protective, e.g high school education, and it is unequally concentrated on the well off, the increase in the effect of this determinant, though a good thing in itself, will increase inequality in childhood ill-health. To put it all together, the change (between 2007 and 2014) in the CI of each measure of childhood ill-health can be written as a weighted sum of three components, namely, the weighted sum of changes in the CI of determinants, the weighted sum of changes in the effects (elasticities) of determinants, and the change in the CI of unobservable determinants.

4.2 Summary of Findings

4.2.1 Results: Paper I

Health shocks had an effect on food and non-food consumption, earned income, health expenditure, and the likelihood of employing self insurance (coping) strategies before and after 2002. This suggests that social protection systems in Zambia, at least at the district level, are inadequate.

In the full sample, injury was associated with a 1.5 percentage point (pp) reduction in non-food consumption share before 2002. However, different income groups were affected differently, with consumption of middle income households (second and third quartile) exhibiting the most significant association. Additionally, the pattern of the effect in these households reflected substitution between food and non-food consumption. In the poorer quartile (second quartile), household cut down food expenditure and increased non-food spending following a health shocks. The opposite was true however for the third quartile.

In the period after 2002, injury was associated with a 6.6% reduction in total food con-

sumption, and a 2.2pp drop in food share. As is the case with the period before 2002, the association between consumption and injury was greatest in middle income households and reflected substitution away from food consumption to non-food consumption.

In general, the associations observed in both periods for any injury in the household was also maintained when injury was restricted to the head of household. However, with fewer number of injuries in the model because of restricting injuries to household heads, standard errors were less tightly estimated.

Turning to health expenditure and income, while injury was associated with a 24.5% increase in household health expenditure in the full sample before 2002, injury almost doubled spending after 2002 (98.1%). The effect on health expenditure was greatest in the two poorest quartile before 2002, but after 2002, all quartiles appear to have been equally affected. For incomes, injury was associated with a 16.2% reduction in earned income in the full sample before 2002, and the effect was the same after 2002, at 16.3%. As was the case with consumption, the association between earned income and health shocks was strongest in middle income households. Injury was not associated with unearned income in the full sample in both periods. However, it was associated with increased unearned income for the poorest households before 2002.

Turning to self insurance strategies, injury was significantly associated with increased likelihood of informal borrowing and selling assets.[4] Injury was associated with a 4.8pp increase in the likelihood of a household borrowing money from informal sources, and this effect was strong for households in all income groups except those in the richest quartile. Similarly, injury was associated with an increased likelihood of selling assets by 5pp in the full sample. Although large in all households, the effect of injury on selling assets was only significant in households belonging to the two richest quartiles. Injury was not associated with formal borrowing, except in the richest quartile.

4.2.2 Results: Paper II

Results show that both waves of the user fee removal in public health facilities, T1 and T2, increased overall utilization of health services with a stronger increase for individuals from lower socioeconomic backgrounds. There is also evidence of individuals shifting care seeking from private to public facilities. In terms of health expenditure, the removal policy signifi-

[4]Note our discussion in Section 4.1.2, that we did not split the analysis into before and after 2002

cantly reduced the proportion of individuals incurring any health spending but did not have any significant effect on conditional spending. In fact, the removal policy exerted an upward pressure on conditional spending and as a result, overall, or unconditional, health expenditure was not affected. These effects were maintained in the long term.

Specifically, the first wave of user fee removals (T1) increased utilization of public health facilities by 10pp in the short term (2006), of which 6.2pp was due to the increase in overall use of health services (uptake effect) and 3.4pp was due to shifting care seeking from private to public facilities (switching effect). The effect strengthened slightly in the long term (2010), to 11.1pp (8pp uptake and 3.2 switching effect). The second wave of user fee removals increased utilization of public health facilities by 13.7pp, of which 11.1pp was due to the uptake effect and 2.7pp was a result of switching, although the switching effect was not statistically significant.

The increase in overall utilization of health services (uptake effect) was greater for individuals whose heads of households had lower educational attainment. In fact, the increase in overall utilization (uptake effect) was only significant for individuals whose household heads either had no education (10.6pp) or who just had primary education (6.2pp). Similarly, overall utilization of health services increased significantly only for individuals from households were the head was not employed. Although, the effect for individuals coming from household heads who were either self or formally employed was positive, it was not significant.

On the contrary, switching from private to public facilities was greater for individuals whose household heads had higher levels of education, although not statistically significant. The effect was only significant for primary education (2.2pp), although lower in magnitude compared to switching effect of secondary and college (3.4pp and 3.1 pp, respectively). Similarly, the removal policy shifted care seeking from private to public facilities for individuals whose household heads were formally employed (4.5pp) or farmers (7.7pp). There was not switching for the unemployed and self employed.

Turning to health expenditure, the first wave of removals reduced the likelihood of incurring any health expenditure by 26.6pp in the short term, but the effect was smaller in the long term (19.3pp). The second wave of removals reduced the likelihood of incurring any health spending by 8.7pp. Both waves of removals increased conditional health spending, although these effects were not statistically significant. On the overall thus, both waves of the removal policy did not affect unconditional health expenditure.

4.2.3 Results: Paper III

In general, the neighborhood/community in which a child lived was important in explaining childhood stunting and fever, with stronger effects for fever. The determinants that were associated with lower likelihood of childhood stunting were wealth; mother's height, weight, education level, and age; household size; and child's birth-weight and age. Stunting was more likely with longer duration of breastfeeding, higher birth order of the child, the child being male, and higher number of children under five years in the household. Similarly, the determinants associated with lower likelihood of childhood fever were wealth; mother's education level and age; higher number of children who are under five years old in the household; and child's age. On the other hand, mother being employed, longer duration of breastfeeding, child being male, and household size, were associated with increased likelihood of fever.

Although stunting reduced between 2007 and 2014, it's inequality increased significantly. Fever incidence did not fall but it's inequality also increased. Specifically, while the prevalence of stunting dropped from 45.6% to 40% between 2007 and 2014, the concentration index (CI) of stunting significantly increased from -0.093 to -0.135 suggesting that stunting became more concentrated on the poor. Stunting prevalence reduced for all wealth quartiles except the poorest. Similarly, the CI of fever significantly increased from -0.015 to -0.064. The incidence of fever did not fall. In fact, it increased slightly from 18.4% to 21.6%.

The increase in the inequality of stunting was accounted for by both the increase in the CI (inequality) of determinants (42.5%) and increase in their effects (35%), measured as elasticities. The determinants with the greatest total contribution (increase in CI plus increase in effect) were mother's height and weight (37%), wealth (32%), birth order (27%), facility delivery (26%), duration of breastfeeding (13%), and maternal education (9%). For some determinants, e.g mother's height, weight, and birth order of the child, the two mechanisms (increase in CI and increase in effect) reinforced each other to drive inequality of stunting up. Although some determinants, such as wealth, maternal education, and facility deliveries, experienced a slight reduction in measured inequality, which could have reduced inequality in stunting, the effect of these determinants increased. Since these determinants were still unequally concentrated on the well off, the increase in the effect of these determinants disproportionately benefited the well off. In other words, the reduction in inequality of these determinants was too small to ensure that the increase in their effect on stunting is equally shared between the poor and the well off. Given that longer duration of breastfeeding was associated with a high

likelihood of stunting, it's contribution to the increase in inequality was due to the fact that it become more unequally concentrated on the poor.

For fever, almost all the increase (86%) in the inequality of fever was accounted for by the increase in the effects of determinants of fever, while the other 14% being due to unobservable determinants. The determinants with the greatest total contribution (increase in CI plus increase in effect) to the increase in inequality of fever were wealth (99%), mother's education (32%), birth order (24%), and duration of breastfeeding (16%).[5] The large contribution of wealth was due to the substantial strengthening of the effect of wealth on fever in 2014, and due to the highly pro-rich distribution of wealth, most of this benefit accrued to the better off. This drove inequality in fever up. The same can be said of maternal education. Moreover, even if almost all the increase in inequality of fever was accounted for by the increase in the effects of determinants, some determinants' contributions were both due to the change in their effects and change in their concentration indices. This can be said of birth order whose contribution to the increase in inequality of fever was due to the two mechanisms reinforcing each other.

[5] Note that the sum of the contribution of these determinants is over 100%. This is because some determinants worked to reduce inequality and hence had a negative percentage contribution. We reported these determinants in the respective tables in the paper. We do not present them here because we are interested in what contributed to the increase in inequality and not what caused inequality not to increase as much as it would have increased

Table 4.1: Summary of Methods

Paper	Data Sources	Empirical Approach
Paper I	Nationally representative household survey data from the 1996, 1998, 2004 and 2006 waves of the living conditions monitoring survey (LCMS). Individuals were asked whether they suffered a sickness or injury in the two weeks prior to the survey. Expenditure on food, non food, and health is based on a one month reference period. For self insurance (or coping), individuals were asked to chose, among the alternatives provided, which strategies they employ in times of hardships	Used injury as a measure of health shock, controlled for cohort, regional, and regional by time effects in additional to demographic and socioeconomic covariates. Assessed the effect of health shocks on: 1. the components of consumption (food and non-food), income (earned and unearned) and coping strategies (selling assets, formal, and informal borrowing,). For each of the three systems of equation, i.e consumption, income, and coping, fitted a Seemingly Unrelated Regression (SUR) separately for the period before and after 2002 [a] 2. household health expenditure by fitting a generalized linear Poisson regression model on medical spending separately for the period before and after 2002.
Paper II	Nationally representative household survey data from the 1998, 2002, 2004, 2006, and 2010 waves of the LCMS. Individuals were asked on whether they utilized any health services in the two weeks prior to the survey and the type of facility type (e.g government clinic or private). All sick individuals were also asked on the health spending they incurred in the two weeks prior to the survey.	Used difference-in-difference (diff-in-diff) regression models taking advantage of the natural experiment provided by the stepwise removal of user fees in public health facilities. Estimated the effect of the removal policy on: 1. overall utilization of health services and switching between private and public facilities. The effect on utilization applied linear diff-in-diff models. The short term (2006) and long term effects (2010) were estimated. Heterogeneity in the utilization response by education and occupation of the head of household was also investigated. 2. household health expenditure. A two part model (TPM) was used were, in the first part, the dependent variable takes the value of 0 or 1 depending on whether somebody incurred any health expenditure. The second part looked at health spending on condition that it was positive, and hence the dependent variable was conditional spending. A probit diff-in-diff model was used for the first part and a generalized linear Gamma regression model for the second part. Marginal effects were reported and standard errors computed using a bootstrap procedure with 1,000 replications. The first and second part of the TPM were combined to assess the effect of the policy on unconditional health spending. .
Paper III	Nationally representative household survey data from the 2007 and 2014 waves of the demographic and health survey (DHS). Height measures were taken by trained enumerators for all children under the age of five. Using the 2006 WHO growth standards, we classified a child as being stunted if their height-for-age was less than two standard deviations of the reference population. Fever was collected in the survey by asking the parents if their child had a fever in the two weeks prior to the survey	1. Used multilevel linear regression models to assess the determinants of childhood ill-health (stunting and fever) and whether the community the child lived was important in explaining childhood ill-health. 2. The Wagstaff corrected Concentration Index was used to quantify inequality in childhood ill-health in 2007 and 2014. The change in the concentration indices and their standard errors were computed. Given the multiple stages involved in the computation, standard errors were estimated using a bootstrap procedure with 1,000 replication. 3. The change in the CI over the period 2007–2014 was decomposed using the Oaxaca type decomposition to assess which determinants had the greatest total contribution to the change in inequality. For each determinant, the change in inequality was decomposed into change in the effect of the determinant and change in the CI of the determinant.

[a] Data for coping strategies was missing for 2004. To allow for controlling of cohort effects, we did not split the analysis to pre- and post-2002.

Chapter 5

Discussion and Conclusion

5.1 Discussion

5.1.1 Discussion of Main Findings

Would increased social protection improve social welfare?

We have examined the effect of health shocks on consumption, assessed the extent to which households are protected from increased health spending and reduced labor income when hit by health shocks, and investigated the extent to which households may rely on informal self insurance mechanisms in the absence of formal social protection. Now the question arises; do our findings suggest that increased social protection in Zambia would increase social welfare? This question is key yet daunting because determining the social welfare impact of social protection rests on comparing the costs and benefits of social protection programs. The economic costs of social protection includes the direct and indirect costs of public funds used to reduce risk and insuring them against its effects.

On the other hand, the extent of the drop in consumption, a key indicator of household welfare, gives a direct measure of the welfare benefit of increased social protection [84]. Another way to assess welfare benefit is to examine the value that households place in preventing a drop in their income or consumption, which is indicated by the extent to which they employ self insurance strategies following a shock. This study focused on the welfare benefit, and not costs, of social protection.

Traditionally, the benefit of social protection has been viewed to be large, the larger the changes in consumption induced by shock related income fluctuations [84]. However, it is

important to state that empirically deducing the welfare benefit of social protection from the consumption response may be difficult since consumption may move for other reasons even when households are protected. The strength of our study is that we also examined the pathways through which health shocks may affect consumption, which are health expenditure and earned income. The results showed substantial income and health expenditure risk giving us confidence that the consumption response we observed may have been due to health shocks. As such, our study suggests that increased social protection would improve households welfare in Zambia. Nonetheless, two issues arise on the extent to which social protection may improve household welfare, given the observed consumption changes. First, even if the effect on consumption was practically large, it was not as much as what we would expect given the large effects on earned income and medical spending. Second, consumption of the poorest quartile was not affected at all, suggesting that additional social protection measures for the poorest 25% may not improve welfare. However, these two conclusions may lack the appreciation that consumption fluctuations following a shock depend on level of household risk aversion [85].

Households in high poverty situations, as is the case in Zambia, may be very risk averse, and this aversion is worsened by the relatively high share of shocks they experience compared to well-off households [117]. This means that they may do all they can, even using very costly self insurance strategies, to prevent consumption fluctuations. Our results suggest that this may be the case for Zambia. We find that households employ self insurance strategies such as informal borrowing and selling assets. Informal borrowing is normally associated with very high monthly interest rates, often as much as 100%. Such inefficient self insurance mechanisms may contribute to the intergenerational transmission of poverty. Thus, a more complete measure of the value of social protection accounts for risk aversion. It is the product of the percentage drop in consumption induced by the shock ($\%\Delta c$) and the coefficient of relative risk aversion (ρ) [85, 118]. This measure implies that even if consumption does not change much following a shock, welfare benefits of social protection may still be large if the coefficient of relative risk aversion is high. Even in developed countries, consumption may be found to be unresponsive to health shocks if households respond by using self insurance strategies such as within household labor substitution to compensate for lost incomes. In this case, as an alternative to using the product of the coefficient of relative risk aversion and consumption change, a number of studies have used spousal labor responses, in terms of

earnings, as a measure of the value of social protection [119].

While our results suggests that social protection will improve household welfare, we are not able to determine whether or not social welfare will improve since we did not look at the economic cost of social protection.

Does providing free health care protect household from health risk and its effects?

The government decided to provide free health care to the population by removing user fees. As a social assistance program, providing free health care may reduce health risk and provide insurance against its effects. The question is whether these objectives were met. Although we were not able to say whether or not the policy reduced health risk since we did not investigate its impact on health, we were able to examine the effect of the policy on utilization of health services, a key pathway through which health may be impacted. The human capital theory, as discussed in Section 3.2.1, predicts that increased investment in health, of which utilization is part, leads to health improvements. Overall utilization increased, especially for individuals from low socioeconomic backgrounds. However, whether or not this translated into reduced health risk depends on many other things.

Among these things, the quality and completeness of freely provided health services is key and it is determined by the adequacy of funding to the health sector. The concern however is that there was no corresponding increase in general government health expenditure (GGHE) after 2006 to meet the extra demand arising from increased utilization of health services (Figure 2.3, Section 2.3.2). Drug stock-outs after the removal of user fees have also been reported [67]. It is thus unclear whether the observed increase in overall utilization could have been accompanied by reduced health risk.

What is clear from our study is that the user fee removal policy was not successful in providing insurance against health expenditure risk. The response of health expenditure to the removal policy may be an important signal of completeness and quality of health services, as well as whether or not health could be impacted. A field experiment of the user fee removal that resulted in substantial reduction in household health spending in Ghana also improved health for some sections of the beneficiaries [120]. In South Africa, were the contribution of households to total health spending is 6% compared to 30% in Zambia (Figure 1.3, Section 1.2.3), the user fee removal policy improved health [121].

Our belief is that improving health and providing insurance against health expenditure risk

depends on the adequacy of health financing. The Zambian health system faces severe financing challenges. On average, general government health expenditure (GGHE) as a percentage of general government expenditure (GGE) has fallen below the abuja target of 15%, despite the huge share of external financing. Moreover, over-reliance on external financing raises concerns about sustainability of health care financing in Zambia. Alternative methods of financing that are being explored, such as social health insurance, should carefully look at the adequacy of finances raised and the extent to which household financial risk will be provided.

Can social protection policy help in curbing the increase in the concentration of health risk among poor children?

We have found that childhood health risk became more concentrated on children from poor families over the 2007–2014 period. Two mechanisms explained this outcome. First, determinants of good childhood health such as income (wealth), facility deliveries, maternal education, mother's height and weight, etc, either remained or become more concentrated on wealthier households while, at the same time, risk factors of childhood health, such as higher birth order and extended breastfeeding duration become more concentrated on poorer household. These finding are consistent with the human capital model of childhood capability formation which predicts that parental capabilities and the environment are important drivers of inequality in childhood capabilities (one of which is health). These findings naturally leads us to the conclusion that the bulk of childhood health inequalities are unjust, they are inequitable.

The second mechanism contributing to the increase in inequality was the strengthening of the sensitivity of childhood health to it's determinants. This finding is consistent with the life cycle model of childhood capability formation which says that inequalities in childhood capabilities, such as health, is not only due to differences in levels of parental capabilities but also the effects of these capabilities (Section 3.2.1). Although, this increase in the effect of determinants (what can be viewed as capabilities) is good–at least for the determinants of good childhood health, the benefits of this strengthening accrued more to children from wealthier households, since health improving determinants were unequally concentrated on them to begin with; this is a classical efficiency-equity trade-off. If so, what are the implications for policy?

We envision two possible pathways of halting the increasing child health inequality corresponding to the two mechanisms that generated them. The first policy intervention is to focus

on reducing inequalities in determinants of good childhood health,such as income (wealth), facility deliveries, maternal education, mother's height and weight, etc. This can involve both short and long term measures. In the short to medium term, cash transfer programs can be used to improve incomes of poor households who have under five children. In this vein, the government may consider reintroducing the child grant component of the social cash transfer (SCT) model which was piloted, showed positive outcomes, but dropped in the scale up of SCT. Alternatively, one of the criteria in the current SCT inclusive model (IM) could be "being among the poor households with children under the age of 5". Cash transfer to poor families with children may increase parental capabilities such as incomes. In the long term, there should be a focus on improving educational outcomes of children who face challenges in schooling, especially girls since the are more disadvantaged and yet they are future mothers. There may be need to scale up social protection schemes such as school feeding programs, that attempt to improve school attendance of the most vulnerable. Although the current government policy focus on basic education is critical, social protection programs, and social spending for education in general, should extend to higher levels of schooling,. Our findings show that it is higher levels of maternal education that are most associated with childhood good health. It is also higher levels of education that result in higher parental incomes.

The second pathway of curbing the increasing childhood health inequalities is to reduce the effect of determinants that are associated with poor childhood health. Results show that longer duration of breastfeeding is associated with poor childhood health and it is concentrated on the poor. Although controversial in the literature, the mechanism through which extended breastfeeding leads to poor childhood health appears to be through inadequate complementary feeding. Current programs such as complementary feeding counseling and support that provide feeding support to malnourished children may need to be scaled up. To do this, it may be necessary to shift this program from being clinic/facility based to population based under, for example the Food Security pack (FSP) scheme. Other social protection programs such as the social cash transfer may also help in eliminating the mediating effect of insufficient complementary feeding in the association between childhood ill-health and breastfeeding duration.

With regard to birth order, it has been found that a large share of differences in childhood capabilities (such as education and health) among siblings are due to the fact that parents tend to invest less in higher birth order children (e.g last borns) [122] . Our findings of birth order being a driver of the increase in inequality suggests that wealthier household are more able to

invest equally in all their children than poorer ones. There are possible unobservable parental capabilities, such as cognitive and non-cognitive skills, that may explain this relationship. It is possible for example that wealthier households, despite having the same educational attainment with poorer households, may have received higher quality education which improved their other capabilities. This may suggest the role of not just improving educational attainment for vulnerable household, but ensuring that it is of good quality.

5.1.2 Discussion of Findings in the Literature

Effect of health shocks on household consumption, income, health expenditure and self insurance

There is a rich literature on assessing the effect of health shocks in developing countries [5, 7, 8, 29, 86–96, 112, 123]. This literature has applied a wide variety of health shock measures which include death, illness, self assessed health, Body Mass Index (BMI), injury/accident, and Activities of Daily Living (ADL). The findings are mixed. With regard to the consumption effect, a number of studies find that health shocks adversely affect household consumption [5, 8, 29, 87, 96], while in other studies, this effect is restricted to subsamples or consumption components [7, 29, 92, 93]. Still others do not find any effect [86, 90, 94, 95, 124], and conclude that consumption is well insured.

In contrast to our study which focused on the general population, most studies from subsaharan Africa are based on rural populations or specific areas. For example, Beegle et al. [87] used a 13 year panel from the Kagera region in Tanzania to assess the short and long term effect of adult mortality on household consumption growth. They found that consumption growth is affected in the first five years following death but households seem to recover afterwards. Using the Ethiopian rural household survey, Asfaw and Von Braun [91] found that the movement of a household head from healthy to unhealthy state substantially lowered the quarterly growth rate of total non-food consumption as well as purchased food consumption. In an assessment of multiple shocks in 15 villages of Ethiopia, Dercon et al. [125] found that only two shocks, namely illness and experiencing a drought, significantly affected levels of consumption. On the other hand, Linnemayr [90] found no differences in consumption changes between households affected by HIV mortality and those that were not .

With regard to the pathways through which health shocks may affect consumption, a vast literature has assessed both income and health expenditure risk. An overwhelming number of studies in Asia find that health shocks substantially increase health expenditure [5, 7, 93], with

sickness almost doubling health expenditure [7, 93]. This is consistent with our finding that injury almost doubled health expenditure in the period following structural adjustment reforms. Apart from our study, we did not find any other study that examined the effect of health shocks on household health expenditure in Africa. Perhaps more than any other outcome, studies in Africa have examined the effect of health shocks on household income or labor force participation or production [88, 89]. Most of these studies are based on rural populations, and have primarily focused on mortality. Using a two year panel of Kenyan rural households, Yamano and Jayne [88] found that death of an adult household member substantially reduces crop and non-farm income and these effects did not fade away within three years of death. Mahmoud and Thiele [86] on the other hand used data from rural Zambia and found no evidence of an effect of adult death on household income.

In contrast to studies from Africa, the literature from Asia has investigated income risk in the general population and used a wider set of health shock measures. A number of studies find that health shocks reduce earned income [5, 7, 93, 126]. Wagstaff [5] found that health shocks, such as hospitalizations, did not significantly affect earned income in some specifications. He reasoned that this was consistent with other household members adjusting labor supply when one falls sick. The ability to adjust labor supply may have limited applications in countries, like Zambia, which have strong labor market rigidities and limited economic opportunities.

In terms of self insurance (coping) strategies, Mitra et al. [94] found that consumption insurance is achieved through self insurance mechanisms such as borrowing, selling assets, and decreased education expenditure. Other studies suggests that an increase in within household labor supply is an important means of insuring consumption against health shocks [124] while others point to the importance of access to microcredit and microfinance institutions [8, 127]. Additionally, access to health insurance is found to help in maintaining investment in childhood human capital in the period of adverse health shocks [124]. In general, health shocks are found to substantially increase the likelihood of depleting assets and borrowing [7, 93, 94]. Tran [128] finds that poorer household are less likely to use these coping strategies because of limited resources and hence they recover more slowly when hit by shocks. This corroborates our findings which showed that the likelihood of selling assets or borrowing were lower, the poorer the household.

Effect of removing user fees for health services

There are broadly two strands of literature assessing the impact of user fees on utilization of health services. The first strand focuses on the impact of introducing user fees while the other looks at the effects of removing them. Most studies that have assessed the impact of introducing user fees have compared levels of utilization in a few government health facilities, before and after the introduction of fees. Results show that the introduction of user fees markedly reduced utilization in Burkina Faso [129], Zambia [33], Papua New Guinea[130], and Kenya [131, 132]. On the other hand, Mwabu et al. [133] used cross sectional household data from Kenya and found that the introduction of user fees dramatically lowered utilization of public facilities. The challenge is that the nature of the policy change could not allow Mwabu et al. [133] to adopt a quasi-experimental study set up, which is important in separating the effect of user fees from other things that may have affected utilization, e.g seasonal diseases like malaria or any other fever. Contrary to these findings, studies in Niger [134] and Cameroon [135] found that the introduction of user fees, when accompanied with quality improvements, actually increase utilization. Van Der Geest et al. [136] found that the reason why most people criticized the introduction of user fees in Zambia was because they believed that it did not improve quality of care, taken to be availability of drugs.

This literature however has a number of shortcomings which casts doubt on the validity of findings. The shortcomings identified include use of routine/facility data–widely perceived to be of poor quality, statistical methods deemed inadequate to account for possible sources of bias, and use of a few observational points or health facilities [137]. The quality of routine data in most developing countries is regarded as very low because of both genuine challenges in ensuring data quality and systematic falsification [138–140].

The second strand of literature assesses the impact on utilization of removing user fees, as opposed to introducing them. Again using routine/facility level data, the removal policy in Niger [68] and the first wave of the removal policy in Zambia [67, 68] was found to have increased utilization of public health facilities. These studies were however not able to assess the impact of the removal policy on overall utilization of health services, and not just that of public facilities. Moreover, the nature of the data–facility records–did not permit an analysis of how different socioeconomic groups may have been affected by the removal policy, an important aspect that this thesis has explored. In Uganda, studies using household level data also find increases in utilization [72, 97], but, unlike the case of Zambia, the nature of the

policy change, where user fees were removed countrywide, did not enable these studies to tease out the effect of user fee removal from other concurrent factors.

With regard to protection from health expenditure risk, there is a near absence of studies that attempt to assess the effect of the removal policies on health expenditure except for studies from Uganda [72, 97]. This is partly justified because such an analysis requires household level data, yet most studies use facility level data. To this end, there has been a gap in the literature–which we have attempted to fill– because the other objective of the user fee removal policy, apart from reducing health risk, via increases in utilization, is to improve financial risk protection. The studies from Uganda, that were able to examine the effect of the removal policy on health expenditure found that the policy did not significantly affect health expenditure [72, 97]. Although these studies were purely descriptive, their findings are corroborated by our results. They explained that the inability of the user fee removal policy to lower medical spending might have due to the fact that households may have being incurring substantial expenses in the private sector due stock-outs of drugs and reagents in public facilities [72, 97]. In Zambia, a qualitative study has suggested that this may have been the case [70].

Changes in childhood health inequalities in the run up to 2015

Substantial improvements in child health in the period leading up to 2015, the target year for the MDG on child health, have been documented [141]. However, a rich literature has shown that children from poorer backgrounds have either experienced worsening health or no improvements in most countries [22, 142, 143]. In other words, inequalities or concentration of poor health among children from poor households have persisted or even worsened, despite some improvements [144]. One particular child health outcome which has experienced worsening inequality is stunting World Health Organization [144]. Bredenkamp et al. [22] assessed 53 countries and found that inequalities in stunting only reduced in 11 countries, increased in 11 others and did not change in the rest of the 31 countries over the 1990–2011 period. In Ethiopia, inequalities in under five deaths increased between 2000 and 2011, although disparities in health intervention coverages were either falling or constant [142]. Similarly for Zambia, while we found rising child health inequality, substantial increases in intervention coverages and accompanying improvements in child health have been documented [145]. The key message that has been highlighted in the literature is that almost all measures of child health, including fever and diarrhea incidence, remain concentrated on the poor globally [146].

The worsening inequalities, despite substantial increases in coverage of health interventions, suggests that halting inequalities is beyond the health sector [147], and one needs to understand what other determinants drive these disparities. Consistent with our study, changes in child health socioeconomic disparities have been attributed to changes in the distribution or inequality of determinants [116, 148, 149], such as wealth/income, maternal education, and unobserved community level effects. These studies also review that the increase in the sensitivity of child health to its determinants, or similarly, the increase in the effect of determinants, has contributed to the increase in child health inequality. The increase in the sensitivity of determinants disproportionately benefit the well off if their distribution favors them in the first place. Inequalities in determinants, such as income, have continued to widen [150], and policy needs to focus on narrowing them if child health inequality is to reduce appreciably.

5.1.3 Discussion of Methodological Limitations

Although we have taken a number of steps to ensure validity of our results, endogeneity problems may still arise. Endogeneity occurs when the covariates/independent variables of interest that are included in the model are correlated with unobservable factors which, due to their unobservability, appear in the error term of the model. In other words, endogeneity is the bias that arises when the included covariates are correlated with error term in the model. This may lead to downward or upward bias depending on whether or not the included covariates are negatively or positively correlated with error term and outcome variable. In most cases, it may be difficult to pin down the overall direction of bias because there are multiple sources of endogeneity, some of which may exert downwards and others upward bias. The problem with endogeneity is that it does not only leads to bias, but this bias does not diminish even with larger samples. In other words, the estimators would be both biased and inconsistent. We are interested in discussing how endogeneity may have affected our findings. We do this for each of the key sources of endogeneity, namely, omitted variables, sample selection bias, non-random missingness, simultaneity, and measurement error. We also highlight some of the steps taken to minimize endogeneity.

Omitted Variable bias (Confounding)

Drawing from empirical studies and our conceptual models, we attempted to include all variables in the data that may be correlated with each other and the outcome at the same time.

As discussed in our conceptual models however, some variables, such as cognitive ability, non-cognitive abilities (time preference, risk attitudes, perseverance, motivation, etc), are not observable. We mentioned, for example, in Paper I, that unobservables such as risk attitudes, are likely to be correlated with both the likelihood of suffering injury and the dependent variables (outcomes), e.g income. If one has access to panel data, these unobservables could be differenced away because they are assumed to be fixed in the short to medium term. To control for some of these unobservables in the absence of panel data, we included a number of fixed effects, e.g birth cohort, regional, and time effects, in the regression models of paper I. We also included a rich set of covariates. In paper II, the difference-in-difference estimator takes care of these fixed effects and we discussed a number of robustness checks such as estimating the models with and without covariates. The fact that findings were broadly unchanged in models with and without covariates reduces the likelihood that results may have been driven by unobservables. Importantly, we could not reject the hypothesis that control and treated units followed common trends, a key identifying assumption of the difference-in-difference model. In paper III, we opted to use a survey, the demographic and health survey (DHS), which has a richer set of covariates, especially health, than a widely used alternative, the living standards monitoring survey (LSMS). Failure to include a rich set of covariates may lead to overestimating the contribution of socioeconomic variables on child health inequality.

Sample selection bias

Sample selectivity bias occurs when some observations in the outcome/dependent variable are missing systematically, e.g for a sub-populations, but not for covariates/independent variables. Sample selectivity may induce a downward bias on our estimators. This is potentially a problem for income in paper I where a significant fraction of the population has zero earned income. Earned income was captured for a period of 1 month prior to the survey. However, for many households, farming is the only sources of income, which only comes periodically [6]. The extent of bias in our estimator depends on whether or not we view the estimated effect to be applicable to non-farm earned income only or a broader measure of earned income (both farm and non-farm). In the former case, sample selectivity issues may not arise. However, if we say that we estimated the effect of injury on broader earned income, then our estimator is unbiased only if the unobservable factors that determined occupational choice (farming, business, formal employment, etc), such as cognitive abilities, are independent of the unobservables that

determined earned income. Such an assumption may be implausible. However, in developing countries were there are significant inequalities of opportunities, selection to occupations may be due to exogenous factors such as kinship rather than unobservables, e.g cognitive abilities. Once individuals sort into occupations, earnings may be determined by factors such as cognitive and non-cognitive abilities, as opposed to kinship. This is however still subjective to debate. To be clear thus, if our objective is to look at the effect of injury on broader earned income regardless of occupation, we have potentially underestimated the effect.[1]

Since all our papers are based on survey data, sample selection bias may also arise due to non-coverage error [152], a form of non-sampling error where a portion of the population has a zero probability of being included in the sample. Unlike the preceding case were a section of the population has missing data on the outcome variable only, data is missing for all variables in this case. [2] This problem may be minimal in this thesis as the sample frames used by the central statistical office (CSO) in all the surveys were based on the census listing of households and a complex survey design was used to ensure that all individuals have a non-zero probability of being included. This sampling procedure however results in unequal selection probabilities and clustering which affects standard errors. Hence, we accounted for complex survey design in all our papers.

Another form of non-sampling error which may lead to sample selectivity is when data are not collected on a significant portion of sampled households due to a number of practical challenges. However, this problem is likely to be minimal because response rates are very high in both LCMS and DHS data, which is in part attributed to vigorous training and supervision of field staff [153].

Non-random missingness

Missing values for independent variables can significantly reduce the sample size on which the model is estimated. If the sample becomes too small, standard errors may be imprecisely estimated, leading to a type two error, a case were you fail to detect the effect too often even when it exists. The concern of endogeneity only arises when values of the independent variables are not missing at random. In this instance, bias may arise if the unobservables that determine

[1] The size of the potential underestimation can be shown to be product of the inverse mills ratio and correlation between the error term in occupation choice equation and the earnings equation. A bivariate distribution for the two error terms is assumed. See discussion in Heckman [151]

[2] Thus, one cannot compute the inverse mills ratio which makes this form of bias difficult to address

outcomes are correlated with whether or not a value of the independent variable is observed. Although missingness was not a big problem in our case due to the type of surveys used, we assessed whether it was reasonable to assume that values of some independent variables were missing at random. To do this, we defined a dummy variable equal to 1 if a household had missing data on the independent variable of interest and zero otherwise, we then ran a linear probability regression model of this dummy variable on the each outcome variable, first without other independent variables and then controlling for other variables. Overall, we found no evidence that any missingness was systematically related to the outcome variables.

Simultaneity Bias

In a regression analysis, we model an outcome as being determined by a covariate. The covariate is assumed to be exogenous, or affected by things outside the model, and hence sometimes called an independent variable. When there is simultaneity however, the covariate is also determined by the outcome. In other words, the outcome and covariate are jointly determined, possibly by other things which we put in the error term as unobservables. This induces an endogeneity problem because the unobservables that determine the outcome, which are in the error term, also determine the covariate. While simultaneity is a problem in both panel and cross-sectional studies, it may be more likely in cross-sectional studies because there is no way of telling that the covariate occurred first, and hence caused the outcome.

We can not rule out simultaneity. In paper I, the effect of health shocks on health expenditure may be biased if households who suffer a health shock had higher health expenditure even before the shock, probably because they had lower stocks of health, e.g chronic conditions, as predicted by the human capital model in Section 3.2.1. In this case, chronic conditions independently increase health expenditure relative to those who do not have them. At the same time, the arrival of the health shock, made more likely by the chronic condition increases health spending. However, such feedback effects are unlikely in our case because our measure of health shocks, injury, as opposed to sickness, is not likely to depend on whether or not somebody has a chronic condition.

There is also a possibility of simultaneity between injury and consumption/income outcomes in paper I. These outcomes may affect the likelihood of injury either positively or negatively. Our results suggest that if this reverse causality is present, then for any two households that are the same in all observables ways, such as occupation, education, unobserved regional,

cohort and time effects, residence, etc, but differ only in terms of consumption and earned income, a household with a lower value of consumption and earned income is more likely to experience injury. It is difficult find a reason why lower earned income would result in more injuries for observationally identical households. One possibility could be that incomes affect travel habits, e.g type of transport, and time spent traveling, and hence road injuries. However, socioeconomic characteristics such as education, occupation, residence etc, may play a much more influential role in governing such habits. Yet still, reverse causality cannot be ruled out.

Measurement error

Another possible source of endogeneity in our study is measurement error in outcomes and covariates. In household survey data, like ours, inaccurate information may be captured depending on, among other things, the order of the questions and responses, which may, for example, lead to starting point bias in selecting responses; the way in which an interviewer asks questions; whether or not the respondents wants to give an answer that is socially desirable; the presence of another family member during the interview; the capacity of the respond to remember (recall) events or expenditure/income, which in turn depends on the length of the reference period; telescoping, in which case events are reported to be within the reference period when they actually occurred outside that period; the level of disaggregation used to collect expenditure/income; length of the questionnaire, which may lead to respondent fatigue; and data entry errors. [3]

This thesis was based on secondary data which makes it hard to assess the extent of these problems. However, the data are from household surveys whose data collection methodology has been tested for decades and designed to minimize some of these errors. For example, we used self reported health shocks (illness, injury, and fever) in all papers which have a reference period of two weeks. Short reference periods, of less than a month, for self reported health status, have been found to be suitable in capturing acute health shocks [155].

Nevertheless, many of the problems mentioned above may still be present despite the level of reliability of these surveys. This makes it difficult to rule out measurement error, which for dummy variables, is termed misclassification. Misclassification in independent variables may lead to downward bias, technically termed attenuation bias [156], if the measurement error–unobservable factors that cause one to make systematic mistakes-e.g forgetting, is correlated

[3] See Kasprzyk [154] for a detailed discussion of the sources of measurement error

with the misclassified independent variable–in our case injury or sickness. In the presence of attenuation, we may have underestimated the effect of health shocks, or the removal of user fees. If measurement error is present in our study but is not correlated with the mis-measured injury or sickness, then our estimators are still consistent. However, the standard errors would be larger, implying that even though our estimated effects may be fine, we may be saying that these effects are insignificant when in fact not.

Measurement error in outcomes variables, which in our case includes expenditure, income, utilization, childhood height, and fever, have traditionally been considered a less serious problem [157]. Measurement errors in expenditure and income information is very common. Normally, there would be problems remembering expenditures, especially when the reference period is too long or if items, for any expenditure or income category, are not disaggregated enough. Less disaggregated expenditure, for example, have been found to yield much lower expenditures [158–160]. Our data is based on very disaggregated expenditure and income items which may minimize measurement error. Moreover, the reference period for income and expenditure variables we used is one month. It has been established that one month reference periods for incomes and expenditures perform very well [161].

Importantly, if measurement error in the outcome variable is present, it generally does not cause the estimators to be inconsistent, although the standard errors are inflated. An exception is when measurement error in the outcome variable is correlated with the independent variables. In this case, the estimator would be biased. This situation would imply that households who experience injury would consistently give wrong information on expenditures and incomes compared to households who did not. It is difficult to imagine why or whether this is possible.

Seasonality

Seasonality occurs in the data when a pattern repeats itself at a particular time period or season, although it may vary in size over time [162]. Assessing seasonality in the data is important because surveys conducted in different seasons of the year may show changes in outcomes that are purely due to weather patterns or major calender events such as new year celebration or Christmas. This is especially important if the time dimension of the data is long.

In paper I, we had two sub-samples, pre-2002 (LCMS 1996 and 1998) and post-2002 (LCMS 2004 and 2006). Starting with the post-2002 subsample, both the 2004 and 2006

surveys were conducted in December, which means seasonality is not a problem. For the pre-2002 subsample, the 1996 survey was conducted from September to November, just before the rainy season, while the 1998 survey was conducted in November and December during the rainy. Since the surveys are very close, and the periods are overlapping, seasonality may not be problem. To be sure, we would need to include a dummy that captures differences in the time when each survey was conducted, however, the overlapping periods makes this impractical.

For paper II, since all outcomes were conditional on incidence of reporting illness/injury, we are concerned with seasonality in the incidence of reporting illness/injury and this will typically reflect in disparate difference across surveys. It turns out that there were no substantial differences in the incidence of reporting illness/injury across different years in all surveys we used (LCMS 1998, 2004, 2006, and 2010) hugely because they were conducted in the same period, rainy season, except for the 2002 survey, which covered the whole year. The incidence of sickness/injury were 9%, 12% , 8%, 7%, and 11% in 1998, 2002, 2004, 2006, and 2010, respectively. Perhaps an important problem would be a case were treated and control areas were surveyed at different times raising differential likelihood of reporting sickness. We conducted this check in the paper and found that there was no differences in likelihood of reporting sickness or injury between treated and control. Thus, the problem of seasonality is likely to be minimal.

In paper III, we may not worry about seasonality for stunting because we do not expect heights to exhibit disparate changes within any given year. Seasonality is a concern nonetheless for fever. The 2007 DHS was conducted from April 2007 to October 2007 while the 2014 DHS was conducted from August 2013 to April 2014. The overlapping periods made it hard for us to define seasonal dummies. If the data had the month in which each household was surveyed, we could have included 11 month of year dummies to minimize the possible seasonality. However, such information was not available.

5.2 Concluding Remarks

Households in Zambia, like most of sub-Saharan Africa, live in highly risky environments, were a plethora of shocks such as job loss, health (illness, injury, death), droughts, crop failure etc, threaten their daily survival , which is already at subsistence level. Of all the shocks that household face, health shocks are the most frequent and pose one of the most serious threats

to household welfare. Typically, they may affect a households' capacity to earn income and substantially raise health expenditure. Thus, to protect household from health shocks and their effect, policy has three instruments at its disposal; (1) improve health or reduce the incidence and severity of health shocks, especially among the most vulnerable such as the poor and children (2) provide income protection or insurance against the labor incomes effects of health shocks, and (3) provide free health care or health insurance to protection household from high health expenditures.

The first sub-study of this thesis, Paper I, examined whether or not health shocks had an effect on household welfare, as measured by household consumption. It then moved on to examine possible sources of this effect, was it through reduced labor income or increased health expenditure or both? In so doing, it implicitly evaluated how successful the policies that used instrument two and three could have faired in providing protection. The paper then questioned whether households could have been using some self insurance mechanisms to cope with emanating risk from increased health spending and reduced income. This is important because self insurance mechanisms may have long term effects.

In the second sub-study, paper II, we evaluated the user fee removal policy, which relates to instrument one and three. We evaluated how the policy affected utilization of health services, envisaged to lead to health improvements (instrument one). We also assessed the extent to which the policy may have reduced health expenditure (instrument three).

The third sub-study relates to instrument one; improving health or simply reducing health risk. It first assessed the determinants of childhood ill-health in 2007 and 2014, and questioned whether the scale up of child health interventions during this period was successful in reducing ill-health among children of lower socioeconomic backgrounds. The focus is on how childhood socioeconomic inequality in ill-health changed during this period and what factors could have accounted for this change.

We found that health shocks pose a significant risk on household welfare as measured by consumptions. Health shocks are associated with substantial increases in health expenditure and reductions in earned income. The effect appeared to be greater in middle incomes households. We also found that the user fee removal policy did not affect health expenditure in any significant way. To cope with the increased risk, households engaged in self insurance mechanism that include informal borrowing and selling assets, potentially perpetuating poverty and contributing to the intergenerational transmission of poverty.

Although the user fee removal policy had the encouraging result of increasing utilization of health service, especial among individuals of low socioeconomic background, findings on health expenditures suggest that more needs to be done to protect households from health expenditure risk. Since health expenditure risk is high in all income quartiles, such a policy should seek to cover the whole population. The implication is that the proposed social health insurance, which will only cover the formal sector, although welcome, should be quickly scaled up to the informal sectors, who constitute more than 85% of the labor force.

There is also need to scale up income protection policies to protect household from the effects of health shocks on earned income. Given that the formal sector is small, income protection can realistic be scaled up using social assistance mechanism such as cash transfers. Currently, these schemes focus on the poorest of the poor. Our findings show that middle income households remain the most affected by health shocks. It is important to realize that most middle income household in Zambia are actually poor. This is because they are in the bottom 75% of per capita consumption, implying that the majority are among the 60% of the population that live in poverty. Hence, current social assistance programs, such as the social cash transfer scheme, may need to improve coverage beyond the poorest of the poor.

Childhood ill health and health shocks over the 2007–2014 period became more concentrated on children from poorer households. Our results suggest that halting the increases in childhood ill health inequality depends on improving levels of, and reducing inequality in, determinants such facilities deliveries, wealth, maternal education, maternal nutrition, complementary feeding (for breastfed children), and child care (related to birth order effects). Social protection policies thus should not just be restricted to scaling up social cash transfer and health programs. Attention also needs to be paid on education, especially for girl children who are normally disadvantaged, and yet they are future mothers.

Bibliography

[1] Robert M. Townsend. Consumption insurance: An evaluation of risk-bearing systems in low-income economies. *Journal of Economic Perspectives*, 9(3):83–102, 1995. URL http://www.aeaweb.org/articles.php?doi=10.1257/jep.9.3.83. 1.1

[2] Rasmus Heltberg and Niels Lund. Shocks, coping, and outcomes for Pakistan's poor: Health risks predominate. *The Journal of Development Studies*, 45(6):889–910, 2009. doi: 10.1080/00220380902802214. URL http://dx.doi.org/10.1080/00220380902802214. 1.1

[3] Adam Wagstaff and Magnus Lindelow. Are health shocks different? evidence from a multishock survey in Laos. *Health Economics*, 23(6):706–718, 2014. URL http://dx.doi.org/10.1002/hec.2944. 1.1

[4] Pilar García-Gómez, Hans van Kippersluis, Owen O'Donnell, and Eddy van Doorslaer. Long-term and spillover effects of health shocks on employment and income. *Journal of Human Resources*, 48(4):873–909, 2013. doi: 10.3368/jhr.48.4.873. URL http://jhr.uwpress.org/content/48/4/873.abstract. 1.1

[5] Adam Wagstaff. The economic consequences of health shocks: Evidence from Vietnam. *Journal of Health Economics*, 26(1):82–100, 2007. URL http://www.sciencedirect.com/science/article/pii/S0167629606000737. 1.1, 3.1.3, 3.2.2, 4.1.2, 5.1.2

[6] FinMark Trust and African Heights. FinScope Zambia 2009: Top Line Findings. Survey Report 42, Government of the Republic of Zambia, Financial Sector Development Plan (FSDP) Secretariat., Lusaka, June 2010. 1.1, 5.1.3

[7] Robert Sparrow, Ellen Van de Poel, Gracia Hadiwidjaja, Athia Yumna, Nila Warda, and Asep Suryahadi. Coping with the economic consequences of ill health in Indonesia.

Health Economics, 23(6):719–728, 2014. URL http://dx.doi.org/10.1002/hec.2945. 1.1, 3.1.3, 4.1.2, 5.1.2

[8] Asadul Islam and Pushkar Maitra. Health shocks and consumption smoothing in rural households: Does microcredit have a role to play? *Journal of Development Economics*, 97(2):232–243, 2012. ISSN 0304-3878. doi: http://dx.doi.org/10.1016/j.jdeveco.2011.05.003. URL http://www.sciencedirect.com/science/article/pii/S0304387811000484. 1.1, 3.1.3, 5.1.2

[9] Janet Currie. Healthy, wealthy, and wise: Socioeconomic status, poor health in childhood, and human capital development. *Journal of Economic Literature*, 47(1):87–122, March 2009. doi: 10.1257/jel.47.1.87. URL http://www.aeaweb.org/articles?id=10.1257/jel.47.1.87. 1.1, 1.1

[10] Douglas Almond and Janet Currie. Killing me softly: The fetal origins hypothesis. *Journal of Economic Perspectives*, 25(3):153–72, 2011. doi: 10.1257/jep.25.3.153. URL http://www.aeaweb.org/articles.php?doi=10.1257/jep.25.3.153. 1.1

[11] Harold Alderman, John Hoddinott, and Bill Kinsey. Long term consequences of early childhood malnutrition. *Oxford Economic Papers*, 58(3):450–474, 2006. doi: 10.1093/oep/gpl008. URL http://oep.oxfordjournals.org/content/58/3/450.abstract. 1.1

[12] Douglas S Berkman, Andres G Lescano, Robert H Gilman, Sonia L Lopez, and Maureen M Black. Effects of stunting, diarrhoeal disease, and parasitic infection during infancy on cognition in late childhood: a follow-up study. *The Lancet*, 359(9306):564 – 571, 2002. ISSN 0140-6736. doi: http://dx.doi.org/10.1016/S0140-6736(02)07744-9. URL http://www.sciencedirect.com/science/article/pii/S0140673602077449. 1.1

[13] Damien Échevin. Measuring vulnerability to asset-poverty in sub-saharan africa. *World Development*, 46:211–222, 2013. ISSN 0305-750X. doi: http://dx.doi.org/10.1016/j.worlddev.2013.02.001. URL http://www.sciencedirect.com/science/article/pii/S0305750X13000405. 1

[14] Cesar G Victora, Adam Wagstaff, Joanna Armstrong Schellenberg, Davidson Gwatkin, Mariam Claeson, and Jean-Pierre Habicht. Applying an equity lens to child health and

mortality: more of the same is not enough. *The Lancet*, 362(9379):233–241, 2003. ISSN 0140-6736. doi: http://dx.doi.org/10.1016/S0140-6736(03)13917-7. URL http://www.sciencedirect.com/science/article/pii/S0140673603139177. 1.1

[15] David Cutler and Adriana Lleras-Muney. *Education and Health: Evaluating Theories and Evidence*. Russell Sage Foundation, New York, 2008. URL https://www.russellsage.org/publications/books/080117.779084. 1.1

[16] Petter Lundborg, Martin Nilsson, and Johan Vikström. Heterogeneity in the impact of health shocks on labour outcomes: evidence from swedish workers. *Oxford Economic Papers*, 67(3):715–739, 2015. doi: 10.1093/oep/gpv034. URL http://oep.oxfordjournals.org/content/67/3/715.abstract. 1.1

[17] Anne Case and Angus S. Deaton. Broken down by work and sex: How our health declines. In David A. Wise, editor, *Analyses in the Economics of Aging*, pages 185–212. University of Chicago Press, August 2005. URL http://www.nber.org/chapters/c10361. 1.1, 3.2.1, 2

[18] Myoung-Hee Kim, Chang yup Kim, Jin-Kyung Park, and Ichiro Kawachi. Is precarious employment damaging to self-rated health? results of propensity score matching methods, using longitudinal data in South Korea. *Social Science & Medicine*, 67(12):1982–1994, 2008. ISSN 0277-9536. doi: http://dx.doi.org/10.1016/j.socscimed.2008.09.051. URL http://www.sciencedirect.com/science/article/pii/S0277953608004589. 1.1

[19] Bastian Ravesteijn, Hans van Kippersluis, and Eddy van Doorslaer. The Wear and Tear on Health: What is the Role of Occupation? Tinbergen Institute Discussion Papers 13-143/V, Tinbergen Institute, September 2013. URL https://ideas.repec.org/p/tin/wpaper/20130143.html. 1.1

[20] World Health Organization. Employment Conditions and Health Inequalities. Technical report, Commission on Social Determinants of Health, World Health Organization, September 2007. 1.1

[21] Wilm Quentin, Olayinka Abosede, Joseph Aka, Patricia Akweongo, Kouassi Dinard, Alex Ezeh, Ramadan Hamed, Patrick Kalambayi Kayembe, Getnet Mitike, Gemini

Mtei, Marguerite Te Bonle, and Leonie Sundmacher. Inequalities in child mortality in ten major african cities. *BMC Medicine*, 12(1):1–12, 2014. ISSN 1741-7015. doi: 10.1186/1741-7015-12-95. URL http://dx.doi.org/10.1186/1741-7015-12-95. 1.1

[22] Caryn Bredenkamp, Leander R Buisman, and Ellen Van de Poel. Persistent inequalities in child undernutrition: evidence from 80 countries, from 1990 to today. *International Journal of Epidemiology*, 2014. doi: 10.1093/ije/dyu075. URL http://ije.oxfordjournals.org/content/early/2014/04/14/ije.dyu075.abstract. 1.1, 5.1.2

[23] David M. Cutler and Adriana Lleras-Muney. Understanding differences in health behaviors by education. *Journal of Health Economics*, 29(1):1–28, 2010. ISSN 0167-6296. doi: http://dx.doi.org/10.1016/j.jhealeco.2009.10.003. URL http://www.sciencedirect.com/science/article/pii/S0167629609001143. 1.1

[24] Ariel Fiszbein, Ravi Kanbur, and Ruslan Yemtsov. Social protection and poverty reduction: Global patterns and some targets. *World Development*, 61:167–177, 2014. ISSN 0305-750X. doi: http://dx.doi.org/10.1016/j.worlddev.2014.04.010. URL http://www.sciencedirect.com/science/article/pii/S0305750X14001181. 1.1

[25] N Loayza. Informality in the process of development and growth. Working paper, World Bank, Development Research Group Macroeconomics and Growth Team, September 2016. 1.1

[26] The United Nations. *Universal Declaration of Human Rights*. dec 1948. 1.1

[27] International Labour Organisation. World Social Protection Report 2014/15 - Building economic recovery, inclusive development and social justice. Technical Report id:5917, International Labour Organisation, June 2014. URL https://ideas.repec.org/p/ess/wpaper/id5917.html. 1.1

[28] A Leive and Ke Xu. Coping with out-of-pocket health payments: empirical evidence from 15 african countries. *Bulletin of the World Health Organization*, 86(11):849–856, 2008. 1.1

[29] J De Weerdt and S Dercon. Risk-sharing networks and insurance against illness. *Journal of Development Economics*, 81(2):337–356., 2006. 1.1, 3.1.3, 5.1.2

[30] Koffi Ekouevi and Aderanti Adepoju. Adjustment, social sectors, and demographic change in sub-Saharan Africa. *Journal of International Development*, 7(1):47–59, 1995. ISSN 1099-1328. doi: 10.1002/jid.3380070104. URL http://dx.doi.org/10.1002/jid.3380070104. 1.1

[31] Howard White. Zambia in the 1990s as a Case of Adjustment in Africa. *African Development Review*, 9(2):56–87, 1997. ISSN 1467-8268. doi: 10.1111/j.1467-8268.1997.tb00155.x. URL http://dx.doi.org/10.1111/j.1467-8268.1997.tb00155.x. 1.1

[32] Kwadwo Konadu-Agyemang. Structural adjustment programs and the perpetuating of poverty and underdevelopment in Africa: Ghana's experience. *Scandinavian Journal of Development Alternatives and Area Studies*, 17(2-3):127–144, June-September 1998. 1.1

[33] Erik Blas and ME Limbambala. User-payment, decentralization and health service utilization in zambia. *Health Policy and Planning*, 16(suppl 2):19–28, 2001. doi: 10.1093/heapol/16.suppl_2.19. URL http://heapol.oxfordjournals.org/content/16/suppl_2/19.abstract. 1.1, 2.2, 3.1.3, 5.1.2

[34] Rene Loewenson. Structural adjustment and health policy in africa. *International Journal of Health Services*, 23(4):717–730, 1993. doi: 10.2190/WBQL-B4JP-K1PP-J7Y3. URL http://joh.sagepub.com/content/23/4/717.abstract. 1.1, 2.2

[35] CSO. Living Conditions Monitoring Survey Report 2002–2003. Survey report, Central Statistical Office, Government of the Republic of Zambia, Lusaka, Zambia, November 2004. 1.1, 2.2, 2.3

[36] Steven Radelet. Aid effectiveness and the millennium development goals. *Center for Global Development Working Paper*, (39), 2004. 1.1

[37] Rob Yates. Universal health care and the removal of user fees. *The Lancet*, 373(9680):2078–2081, 2009. 1.1

[38] World Bank. Poverty & Equity, 2016. Data retrived from the World Bank Poverty & Equity Database on 2016-07-12, http://povertydata.worldbank.org/poverty/home/. 1.2.1, 1.1

[39] World Bank. DISTRIBUTION OF INCOME OR CONSUMPTION BY QUINTILE, 2016. Data retrived from the World Bank Poverty & Equity Database on 2016-07-12, http://povertydata.worldbank.org/poverty/region/SSA. 1.2.1, 2.3

[40] World Health Organization. Adult mortality data by WHO region, 2013. Data retrived from the Global Health Observatory data repository on 2016-07-11, http://apps.who.int/gho/data/view.main.1340?lang=en. 1.2.2

[41] Joint United Nations Programme on HIV/AIDS . FACT SHEET 2016. Data sheet, Joint United Nations Programme on HIV/AIDS (UNAIDS), 2016. 1.2.2

[42] WHO Global Malaria Programme. World Malaria Report 2015. Technical report, World Health Organization, 2015. 1.2.2

[43] World Health Organization. Trends in maternal mortality: 1990 to 2015: estimates by WHO, UNICEF, UNFPA, World Bank Group and the United Nations Population Division. Technical report, WHO, UNICEF, UNFPA, World Bank Group and the United Nations Population Division, 2015. 1.2.2

[44] United Nations Children's Fund . Committing to Child Survival: A Promise Renewed. Technical report, United Nations Children's Fund (UNICEF), September 2015. 1.2.2

[45] United Nations Children's Fund. Undernutrition contributes to nearly half of all deaths in children under 5 and is widespread in Asia and Africa, 2016. Accessed on 2016-07-11, http://data.unicef.org/nutrition/malnutrition.html. 1.2.2, 3.1.3

[46] Juanita A Haagsma, Nicholas Graetz, Ian Bolliger, Mohsen Naghavi, Hideki Higashi, Erin C Mullany, Semaw Ferede Abera, Jerry Puthenpurakal Abraham, Koranteng Adofo, Ubai Alsharif, Emmanuel A Ameh, Walid Ammar, Carl Abelardo T Antonio, Lope H Barrero, Tolesa Bekele, Dipan Bose, Alexandra Brazinova, Ferrán Catalá-López, Lalit Dandona, Rakhi Dandona, Paul I Dargan, Diego De Leo, Louisa Degenhardt, Sarah Derrett, Samath D Dharmaratne, Tim R Driscoll, Leilei Duan, Sergey Petrovich Ermakov, Farshad Farzadfar, Valery L Feigin, Richard C Franklin,

Belinda Gabbe, Richard A Gosselin, Nima Hafezi-Nejad, Randah Ribhi Hamadeh, Martha Hijar, Guoqing Hu, Sudha P Jayaraman, Guohong Jiang, Yousef Saleh Khader, Ejaz Ahmad Khan, Sanjay Krishnaswami, Chanda Kulkarni, Fiona E Lecky, Ricky Leung, Raimundas Lunevicius, Ronan Anthony Lyons, Marek Majdan, Amanda J Mason-Jones, Richard Matzopoulos, Peter A Meaney, Wubegzier Mekonnen, Ted R Miller, Charles N Mock, Rosana E Norman, Ricardo Orozco, Suzanne Polinder, Farshad Pourmalek, Vafa Rahimi-Movaghar, Amany Refaat, David Rojas-Rueda, Nobhojit Roy, David C Schwebel, Amira Shaheen, Saeid Shahraz, Vegard Skirbekk, Kjetil Søreide, Sergey Soshnikov, Dan J Stein, Bryan L Sykes, Karen M Tabb, Awoke Misganaw Temesgen, Eric Yeboah Tenkorang, Alice M Theadom, Bach Xuan Tran, Tommi J Vasankari, Monica S Vavilala, Vasiliy Victorovich Vlassov, Solomon Meseret Woldeyohannes, Paul Yip, Naohiro Yonemoto, Mustafa Z Younis, Chuanhua Yu, Christopher J L Murray, and Theo Vos. The global burden of injury: incidence, mortality, disability-adjusted life years and time trends from the global burden of disease study 2013. *Injury Prevention*, 2015. doi: 10.1136/injuryprev-2015-041616. URL http://injuryprevention.bmj.com/content/early/2015/10/20/injuryprev-2015-041616.abstract. 1.2.2

[47] World Health Organization. Crude death rate by broad cause group, 2000 and 2012 by WHO region, 2012. Data retrived from the Global Health Observatory data repository on 2016-07-11, http://apps.who.int/gho/data/view.main.1340?lang=en. 1.2.2

[48] World Bank. Coverage of SPL Programs, 2016. Data retrived from the "THE ATLAS OF SOCIAL PROTECTION - INDICATORS OF RESILIENCE AND EQUITY (ASPIRE)" on 2016-07-18, http://datatopics.worldbank.org/aspire/country/zambia. 1.2.3, 1.2, 2.3.1

[49] World Health Organization. State of Health Financing in the African Region. Report, World Health Organization, WHO Regional Office for Africa, congo, Brazzaville, January 2013. 1.2.3

[50] World Health Organization. National Health Accounts (NHA) Indicators, 2015. Data retrived for the WHO Global Health Expenditure Database on 2016-07-11, http://

apps.who.int/nha/database/Select/Indicators/en.2016-07-11. 1.3, 1.4, 2.2, 2.3.2, 2.2, 2.3

[51] CSO. LCMS 2015 Summary Report. Survey report, Central Statistical Office, Government of the Republic of Zambia, Lusaka, Zambia, 2016. 2.1, 4

[52] World Bank. Gross National Income (GNI) per capita in Zambia and sub-Saharan Africa, 2016. Data retrived from the World Bank world development indicators on 2016-07-12, http://databank.worldbank.org/data/reports.aspx?source=world-development-indicators. 2.1, 2.1

[53] Jeffrey J. Hawkins. Understanding the failure of imf reform: The zambian case. *World Development*, 19(7):839 – 849, 1991. ISSN 0305-750X. doi: http://dx.doi.org/10.1016/0305-750X(91)90137-7. URL http://www.sciencedirect.com/science/article/pii/0305750X91901377. 2.1

[54] Oliver S Saasa. Policy reforms and structural adjustment in zambia: The case of agriculture and trade. 1996. 2.1, 2.1, 2.2

[55] Neo Simutanyi. The politics of structural adjustment in zambia. *Third World Quarterly*, 17(4):825–839, 1996. ISSN 01436597, 13602241. URL http://www.jstor.org/stable/3993288. 2.1, 2.2

[56] F Masiye, V Seshamani, C Cheelo, C Mphuka, and K. Odegaard. Health care financing in Zambia, a study of the possible policy options for implementation. In C Cheelo, N Manenga, and K Odegaard, editors, *Policy Lessons and Emerging Challenges in the Zambian Health Sector 2000-2005.*, pages 77–109. Department of Economics, UNZA and the Swedish Institute for Health Economics, 2006. 2.2

[57] William Easterly. What did structural adjustment adjust?: The association of policies and growth with repeated imf and world bank adjustment loans. *Journal of development economics*, 76(1):1–22, 2005. 2.2

[58] Neil McCulloch, Bob Baulch, and Milasoa Cherel-Robson. Poverty, inequality and growth in zambia during the 1990s. Discussion paper no. 2001/123, United Nations University WIDER, November 2001. 2.2

[59] Christopher Shaw David M. Todd. The informal sector and Zambia's employment crisis. *The Journal of Modern African Studies*, 18(3):411–425, 1980. ISSN 0022278X, 14697777. URL http://www.jstor.org/stable/160362. 2.2

[60] Central Statistical Office (CSO) [Zambia] and Ministry of Health (MOH) [Zambia] and ICF International. Zambia Demographic and Health Survey 2013-14. Survey report, Lusaka, Zambia, November 2014. 2.2, 2.3.2, 2

[61] CSO. Living Conditions Monitoring Survey Report 2006 and 2010. Survey report, Central Statistical Office, Government of the Republic of Zambia, Lusaka, Zambia, March 2012. 2.3

[62] International Labour Organization. Zambia. Social Protection Expenditure and Performance Review and Social Budget . Technical report, Social Security Department, International Labour Office, Geneva, June 2008. 2.3, 2.3.2

[63] Monika Queisser, Clive Bailey, and John Woodall. Reforming pensions in zambia: An analysis of existing schemes and options for reform. *World Bank policy research working paper*, (1716), 1997. 2.3.1

[64] Government of the Republic of Zambia. Fifth National Development Plan, 2006–2010. Government document, Lusaka, Zambia, December 2006. 2.3.1

[65] Maarten van Klaveren, Kea Tijdens, Melanie Hughie-Williams, and Nuria Ramos Martin. An Overview of Women's Work and Employment in Zambia . Decisions for Life MDG3 Project Country Report 4, University of Amsterdam /Amsterdam Institute for Advanced Labour Studies (AIAS) ., Amsterdam, September 2009. 2.3.1

[66] International Labour Organisation. World Social Protection Report 2014/15 - Building economic recovery, inclusive development and social justice. Report to the government, Social Protection Department; Conditions of Work and Equality Department, Decent Work Team, Eastern and Southern Africa; Country Office, Lusaka, June 2014. 2.3.1

[67] Felix Masiye, Bona M. Chitah, and Diane McIntyre. From targeted exemptions to user fee abolition in health care: Experience from rural Zambia. *Social Science & Medicine*, 71(4):743 –750, 2010. ISSN 0277-9536. doi: http://dx.doi.org/10.

1016/j.socscimed.2010.04.029. URL http://www.sciencedirect.com/science/article/pii/S027795361000359X. 2.3.2, 3.1.3, 5.1.1, 5.1.2

[68] Mylene Lagarde, Helene Barroy, and Natasha Palmer. Assessing the effects of removing user fees in Zambia and Niger. *Journal of Health Services Research & Policy*, 17 (1):30–36, 2012. doi: 10.1258/jhsrp.2011.010166. URL http://hsr.sagepub.com/content/17/1/30.abstract. 2.3.2, 5.1.2

[69] R. Onde. Impact of abolishment of user fees in rural health centres: A case of Chongwe district. Master's thesis, School of Business, Copperbelt University, Kitwe, Zambia, March 2009. 2.3.2

[70] Mary Hadley. Does increase in utilisation rates alone indicate the success of a user fee removal policy? a qualitative case study from Zambia. *Health Policy*, 103(2–3):244–254, 2011. ISSN 0168-8510. doi: http://dx.doi.org/10.1016/j.healthpol.2011.08.009. URL http://www.sciencedirect.com/science/article/pii/S0168851011001801. 2.3.2, 5.1.2

[71] Felix Masiye, Oliver Kaonga, and Joses M Kirigia. Does user fee removal policy provide financial protection from catastrophic health care payments? evidence from Zambia. *PLoS ONE*, 11(1):1–15, 01 2016. doi: 10.1371/journal.pone.0146508. URL http://dx.doi.org/10.1371%2Fjournal.pone.0146508. 2.3.2

[72] Juliet Nabyonga Orem, Frederick Mugisha, Christine Kirunga, Jean Macq, and Bart Criel. Abolition of user fees: the Uganda paradox. *Health Policy and Planning*, 26(suppl 2):ii41–ii51, 2011. doi: 10.1093/heapol/czr065. URL http://heapol.oxfordjournals.org/content/26/suppl_2/ii41.abstract. 2.3.2, 3.1.3, 5.1.2

[73] Ann Danaiya Usher. Key donors to reinstate health funding to zambia. *The Lancet*, 386(9993):519–520, 2015. ISSN 0140-6736. doi: http://dx.doi.org/10.1016/S0140-6736(15)61463-5. URL http://www.sciencedirect.com/science/article/pii/S0140673615614635. 3

[74] Stanfield Michelo. Social cash transfer scale-up for Zambia. *The IPC-IG One-Pager*, 287, 04 2015. 2.3.2

[75] American Institutes for Research. Zambia's Child Grant Program: 24-Month Impact Report. Technical report, Washington DC, December 2013. 2.3.2

[76] American Institutes for Research. Zambia's Multiple Category Targeting Grant: 24-Month Impact Report. Technical report, Ministry of Community Development, Mother and Child Health. Government of the Republic of Zambia, December 2014. 2.3.2

[77] Sudhanshu Handa, Amber Peterman, David Seidenfeld, and Gelson Tembo. Income transfers and maternal health: Evidence from a national randomized social cash transfer program in zambia. *Health Economics*, 25(2):225–236, 2016. ISSN 1099-1050. doi: 10.1002/hec.3136. URL http://dx.doi.org/10.1002/hec.3136. 2.3.2

[78] Rodolfo Beazley and Ludovico Carraro. Assessment of the Zambia Social Protection Expansion Programme Targeting Mechanisms . Technical report, Oxford Policy Management and RuralNet Associates Ltd, September 2013. 2.3.2

[79] Ministry of Community Development, Mother, and Child Health. Social Cash Transfer Fact Sheet / Social Cash Transfer (SCT) Programme, Overview. Government document, Government of the Republic of Zambia (GRZ), Lusaka, Zambia, September 2014. 2.3.2

[80] Nicole M. Mason, T.S. Jayne, and Rhoda Mofya-Mukuka. Zambia's input subsidy programs. *Agricultural Economics*, 44(6):613–628, 2013. ISSN 1574-0862. doi: 10.1111/agec.12077. URL http://dx.doi.org/10.1111/agec.12077. 2.3.2

[81] Ministry of Education, Science, Vocational Training and Early Education. Education for All 2015 National Review . National report, Governement of the Republic of Zambia (GRZ), June 2015. 2.3.2

[82] World Food Programme. Home Grown School Feeding Programme Fact Sheet . Programme summary, Lusaka, Zambia, 2013. 2.3.2

[83] The Centre for Financial Regulation and Inclusion. Zambian private health insurance: optimal regulation and market development. http://cenfri.org/documents/cost, October 2009. Online; accessed 13-May-2015. 2.3.2

[84] Jonathan Morduch. Income smoothing and consumption smoothing. *Journal of Economic Perspectives*, 9(3):103–114, September 1995. doi: 10.1257/jep.9.3.103. URL http://www.aeaweb.org/articles?id=10.1257/jep.9.3.103. 3.1.3, 5.1.1

[85] Raj Chetty and Adam Looney. Consumption smoothing and the welfare consequences of social insurance in developing economies. *Journal of Public Economics*, 90(12):2351–2356, 2006. ISSN 0047-2727. doi: http://dx.doi.org/10.1016/j.jpubeco.2006.07.002. URL http://www.sciencedirect.com/science/article/pii/S004727270600082X. 3.1.3, 5.1.1

[86] Toman Omar Mahmoud and Rainer Thiele. Does prime-age mortality reduce per-capita household income? evidence from rural Zambia. *World Development*, 45:51–62, 2013. ISSN 0305-750X. doi: http://dx.doi.org/10.1016/j.worlddev.2012.09.005. URL http://www.sciencedirect.com/science/article/pii/S0305750X12002197. 3.1.3, 5.1.2

[87] Kathleen Beegle, Joachim De Weerdt, and Stefan Dercon. Adult mortality and consumption growth in the age of HIV/AIDS. *Economic Development and Cultural Change*, 56(2):299–326, 2008. 3.1.3, 5.1.2

[88] Takashi Yamano and T.S Jayne. Measuring the impacts of working-age adult mortality on small-scale farm households in Kenya. *World Development*, 32(1): 91–119, 2004. URL http://www.sciencedirect.com/science/article/pii/S0305750X03001992. 3.1.3, 5.1.2

[89] A Chapoto and T Jayne. Impact of AIDS-related mortality on farm household welfare in Zambia. *Economic Development and Cultural Change*, 56(2):327–374., 2008. 3.1.3, 5.1.2

[90] S. Linnemayr. Consumption smoothing and HIV/AIDS: The case of two communities in South Africa. *Economic Development and Cultural Change*, 58(3):475–506., 2010. 3.1.3, 4.1.2, 5.1.2

[91] A Asfaw and J Von Braun. Is consumption insured against illness? Evidence on vulnerability of households to health shocks in rural Ethiopia. *Economic Development and Cultural Change*, 53(1):115–129, 2004. 3.1.3, 5.1.2

[92] Stefan Dercon and Pramila Krishnan. In sickness and in health: Risk sharing within households in rural Ethiopia. *Journal of Political Economy*, 108(4):688–727, 2000. URL http://www.jstor.org/stable/10.1086/316098. 3.1.3, 5.1.2

[93] Farid Khan, Arjun S. Bedi, and Robert Sparrow. Sickness and death: Economic consequences and coping strategies of the urban poor in Bangladesh. *World Development*, 72:255–266, 2015. ISSN 0305-750X. doi: http://dx.doi.org/10.1016/j.worlddev.2015.03.008. URL http://www.sciencedirect.com/science/article/pii/S0305750X15000674. 3.1.3, 5.1.2

[94] Sophie Mitra, Michael Palmer, Daniel Mont, and Nora Groce. Can households cope with health shocks in vietnam? *Health Economics*, pages n/a–n/a, 2015. ISSN 1099-1050. doi: 10.1002/hec.3196. URL http://dx.doi.org/10.1002/hec.3196. 3.1.3, 5.1.2

[95] Robert M. Townsend. Risk and insurance in village India. *Econometrica*, 62(3):539–591, 1994. URL http://www.jstor.org/stable/2951659. 3.1.3, 3.2.2, 5.1.2

[96] P. Gertler and J. Gruber. Insuring consumption against illness. *American Economic Review*, 92(1):51–76, 2002. 3.1.3, 5.1.2

[97] Ke Xu, David B. Evans, Patrick Kadama, Juliet Nabyonga, Peter Ogwang Ogwal, Pamela Nabukhonzo, and Ana Mylena Aguilar. Understanding the impact of eliminating user fees: Utilization and catastrophic health expenditures in uganda. *Social Science & Medicine*, 62(4):866 – 876, 2006. ISSN 0277-9536. doi: http://dx.doi.org/10.1016/j.socscimed.2005.07.004. URL http://www.sciencedirect.com/science/article/pii/S0277953605003667. 3.1.3, 5.1.2

[98] Gary S. Becker. A theory of the allocation of time. *The Economic Journal*, 75(299):493–517, 1965. ISSN 00130133, 14680297. URL http://www.jstor.org/stable/2228949. 1, 3.2.1

[99] Michael Grossman. Chapter 7 the human capital model. volume 1, Part A of *Handbook of Health Economics*, pages 347–408. Elsevier, 2000. doi: http://dx.doi.org/10.1016/S1574-0064(00)80166-3. URL http://www.sciencedirect.com/science/article/pii/S1574006400801663. 1, 3.2.1

[100] James J. Heckman. The economics, technology, and neuroscience of human capability formation. *Proceedings of the National Academy of Sciences*, 104(33):13250–13255, 2007. doi: 10.1073/pnas.0701362104. URL http://www.pnas.org/content/104/33/13250.abstract. 1, 3.2.1

[101] J Cochrane. A simple test of consumption insurance. *Journal of Political Economy*, 99 (5):957–976., 1991. 1, 3.2.2

[102] Yoram Ben-Porath. The production of human capital and the life cycle of earnings. *Journal of Political Economy*, 75(4):352–365, 1967. ISSN 00223808, 1537534X. URL http://www.jstor.org/stable/1828596. 3.2.1

[103] Jacob A. Mincer. *Schooling, Experience, and Earnings*. Number minc74-1 in NBER Books. National Bureau of Economic Research, Inc, 1974. URL https://ideas.repec.org/b/nbr/nberbk/minc74-1.html. 3.2.1

[104] Gary S. Becker, Michael Grossman, and Kevin M. Murphy. An empirical analysis of cigarette addiction. *The American Economic Review*, 84(3):396–418, 1994. ISSN 00028282. URL http://www.jstor.org/stable/2118059. 3.2.1

[105] Michael Grossman. On the concept of health capital and the demand for health. *Journal of Political Economy*, 80(2):223–255, 1972. ISSN 00223808, 1537534X. URL http://www.jstor.org/stable/1830580. 3.2.1

[106] Maureen L Cropper. Health, investment in health, and occupational choice. *The Journal of Political Economy*, pages 1273–1294, 1977. 3.2.1

[107] Jaana-Marja Muurinen. Demand for health. *Journal of Health Economics*, 1(1):5 –28, 1982. ISSN 0167-6296. doi: http://dx.doi.org/10.1016/0167-6296(82)90019-4. URL http://www.sciencedirect.com/science/article/pii/0167629682900194. 3.2.1

[108] Jaana-Marja Muurinen and Julian Le Grand. The economic analysis of inequalities in health. *Social Science & Medicine*, 20(10):1029 – 1035, 1985. ISSN 0277-9536. doi: http://dx.doi.org/10.1016/0277-9536(85)90259-X. URL http://www.sciencedirect.com/science/article/pii/027795368590259X. 3.2.1

[109] Hans van Kippersluis and Titus J. Galama. Wealth and health behavior: Testing the concept of a health cost. *European Economic Review*, 72:197 – 220, 2014. ISSN 0014-2921. doi: http://dx.doi.org/10.1016/j.euroecorev.2014.10.003. URL http://www.sciencedirect.com/science/article/pii/S0014292114001408. 3.2.1, 3

[110] Pedro Carneiro, Flavio Cunha, and James J Heckman. Interpreting the evidence of family influence on child development. *The Economics of Early Childhood Development: Lessons for Economic Policy*, 2003. 3.2.1

[111] Samuel Bowles, Herbert Gintis, and Melissa Osborne Groves. *Unequal chances: Family background and economic success*. Princeton University Press, 2009. 3.2.1

[112] Anne Moller Dano. Road injuries and long-run effects on income and employment. *Health Economics*, 14(9):955–970, 2005. ISSN 1099-1050. doi: 10.1002/hec.1045. URL http://dx.doi.org/10.1002/hec.1045. 4.1.2, 5.1.2

[113] Arnold Zellner. An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias. *Journal of the American Statistical Association*, 57(298):348–368, 1962. URL http://www.tandfonline.com/doi/abs/10.1080/01621459.1962.10480664. 4.1.2

[114] M Frondel and C Vance. On Interaction Effects: The Case of Heckit and Two-Part Models. *Journal of Economics and Statistics*, 233(1):22 –38, 2013. 4.1.3

[115] Adam Wagstaff. The bounds of the concentration index when the variable of interest is binary, with an application to immunization inequality. *Health Economics*, 14(4):429–432, 2005. ISSN 1099-1050. doi: 10.1002/hec.953. URL http://dx.doi.org/10.1002/hec.953. 4.1.4

[116] Adam Wagstaff, Eddy van Doorslaer, and Naoko Watanabe. On decomposing the causes of health sector inequalities with an application to malnutrition inequalities in vietnam. *Journal of Econometrics*, 112(1):207–223, 2003. ISSN 0304-4076. doi: http://dx.doi.org/10.1016/S0304-4076(02)00161-6. URL http://www.sciencedirect.com/science/article/pii/S0304407602001616. Analysis of data on health: 2. 4.1.4, 5.1.2

[117] Oliver Gloede, Lukas Menkhoff, and Hermann Waibel. Shocks, individual risk attitude, and vulnerability to poverty among rural households in Thailand and Vietnam. *World Development*, 71:54–78, 2015. ISSN 0305-750X. doi: http://dx.doi.org/10.1016/j.worlddev.2013.11.005. URL http://www.sciencedirect.com/science/article/pii/S0305750X13002477. Vulnerability to Poverty in South-East Asia: Drivers, Measurement, Responses, and Policy Issues. 5.1.1

[118] Raj Chetty. A general formula for the optimal level of social insurance. *Journal of Public Economics*, 90(10–11):1879 – 1901, 2006. ISSN 0047-2727. doi: http://dx.doi.org/10.1016/j.jpubeco.2006.01.004. URL http://www.sciencedirect.com/science/article/pii/S0047272706000223. 5.1.1

[119] David Autor, Andreas Ravndal Kostøl, and Magne Mogstad. Disability benefits, consumption insurance, and household labor supply. Working paper, Becker Friendman Institute For Reserach in Economics, The University of Chicago, September 2015. 5.1.1

[120] Timothy Powell-Jackson, Kara Hanson, Christopher J.M. Whitty, and Evelyn K. Ansah. Who benefits from free healthcare? evidence from a randomized experiment in Ghana. *Journal of Development Economics*, 107:305–319, 2014. ISSN 0304-3878. doi: http://dx.doi.org/10.1016/j.jdeveco.2013.11.010. URL http://www.sciencedirect.com/science/article/pii/S0304387813001727. 5.1.1

[121] Shinsuke Tanaka. Does abolishing user fees lead to improved health status? evidence from post-apartheid South Africa. *American Economic Journal: Economic Policy*, 6(3):282–312, 2014. doi: 10.1257/pol.6.3.282. URL http://www.aeaweb.org/articles.php?doi=10.1257/pol.6.3.282. 5.1.1

[122] Ronni Pavan. On the production of skills and the birth order effect. *Journal of Human Resources*, 2015. doi: 10.3368/jhr.51.3.0913-5920R. URL http://jhr.uwpress.org/content/early/2015/11/20/jhr.51.3.0913-5920R.abstract. 5.1.1

[123] M. Mohanan. Causal effects of health shocks on consumption and debt: Quasi-experimental evidence from bus accident injuries. *Review of Economics and Statistics*, 95(2):673–681, 2011. 5.1.2

[124] Kai Liu. Insuring against health shocks: Health insurance and household choices. *Journal of health economics*, 46:16–32, 2016. 5.1.2

[125] Stefan Dercon, John Hoddinott, and Tassew Woldehanna. Shocks and consumption in 15 ethiopian villages, 1999-2004. *Journal of African Economies*, 2005. doi: 10.1093/jae/eji022. URL http://jae.oxfordjournals.org/content/early/2005/08/19/jae.eji022.short. 5.1.2

[126] Maria Eugenia Genoni. Health shocks and consumption smoothing: Evidence from indonesia. *Economic Development and Cultural Change*, 60(3):475?506, 2012. ISSN 00130079, 15392988. URL http://www.jstor.org/stable/10.1086/664019. 5.1.2

[127] Paul Gertler, David I. Levine, and Enrico Moretti. Do microfinance programs help families insure consumption against illness? *Health Economics*, 18(3):257–273, 2009. ISSN 1099-1050. doi: 10.1002/hec.1372. URL http://dx.doi.org/10.1002/hec.1372. 5.1.2

[128] Van Q. Tran. Household's coping strategies and recoveries from shocks in vietnam. *The Quarterly Review of Economics and Finance*, 56:15 –29, 2015. ISSN 1062-9769. doi: http://dx.doi.org/10.1016/j.qref.2014.06.006. URL http://www.sciencedirect.com/science/article/pii/S1062976914000477. 5.1.2

[129] Valéry Ridde. Fees-for-services, cost recovery, and equity in a district of burkina faso operating the bamako initiative. *Bulletin of the World Health Organization*, 81(7):532–538, 2003. 5.1.2

[130] AL Benjamin, P Sapak, and JK Purai. User charges and utilisation of obstetric services in the national capital district, papua new guinea. *Pacific health dialog*, 8(1):38–43, 2001. 5.1.2

[131] J.Karanja Mbugua, Gerald H. Bloom, and Malcolm M. Segall. Impact of user charges on vulnerable groups: The case of kibwezi in rural kenya. *Social Science & Medicine*, 41(6):829–835, 1995. ISSN 0277-9536. doi: http://dx.doi.org/10.1016/0277-9536(94)00400-N. URL http://www.sciencedirect.com/science/article/pii/027795369400400N. 5.1.2

[132] David Collins, Jonathan D Quick, Stephen N Musau, Daniel Kraushaar, and Ibrahim M Hussein. The fall and rise of cost sharing in kenya: the impact of phased implementation. *Health policy and planning*, 11(1):52–63, 1996. 5.1.2

[133] Germano Mwabu, J Mwanzia, and W Liambila. User charges in government health facilities in Kenya: effect on attendance and revenue. *Health Policy and Planning*, 10(2):164–170, 1995. doi: 10.1093/heapol/10.2.164. URL http://heapol.oxfordjournals.org/content/10/2/164.abstract. 5.1.2

[134] FRANCOIS DIOP, ABDO YAZBECK, and RICARDO BITRAN. The impact of alternative cost recovery schemes on access and equity in niger. *Health Policy and Planning*, 10(3):223–240, 1995. doi: 10.1093/heapol/10.3.223. URL http://heapol.oxfordjournals.org/content/10/3/223.abstract. 5.1.2

[135] Jennie I. Litvack and Claude Bodart. User fees plus quality equals improved access to health care: Results of a field experiment in Cameroon. *Social Science & Medicine*, 37 (3):369 –383, 1993. doi: http://dx.doi.org/10.1016/0277-9536(93)90267-8. 5.1.2

[136] Sjaak Van Der Geest, Mubiana Macwan?gi, Jolly Kamwanga, Dennis Mulikelela, Arthur Mazimba, and Mundia Mwangelwa. User fees and drugs: what did the health reforms in zambia achieve? *Health Policy and Planning*, 15(1):59–65, 2000. 5.1.2

[137] Mylene Lagarde and Natasha Palmer. The impact of user fees on health service utilization in low- and middle-income countries: how strong is the evidence? *Bulletin of the World Health Organization*, 86:839 – 848C, 11 2008. ISSN 0042-9686. URL http://www.scielosp.org/scielo.php?script=sci_arttext&pid=S0042-96862008001100013&nrm=iso. 5.1.2

[138] Justin Sandefur and Amanda Glassman. The Political Economy of Bad Data: Evidence from African Survey and Administrative Statistics. *The Journal of Development Studies*, 51(2):116–132, 2015. doi: 10.1080/00220388.2014.968138. 5.1.2

[139] Stephen S Lim, David B Stein, Alexandra Charrow, and Christopher JL Murray. Tracking progress towards universal childhood immunisation and the impact of global initiatives: a systematic analysis of three-dose diphtheria, tetanus, and pertussis immunisation coverage. *The Lancet*, 372(9655):2031–2046, 2008. ISSN 0140-6736. doi: http://dx.doi.org/10.1016/S0140-6736(08)61869-3. URL http://www.sciencedirect.com/science/article/pii/S0140673608618693. 5.1.2

[140] Nava Ashraf, Günther Fink, and David N Weil. Evaluating the effects of large scale health interventions in developing countries: The Zambian Malaria Initiative. In Sebastian Edwards, Simon Johnson, and David N. Weil, editors, *African Successes: Human Capital, Volume 2*. University of Chicago Press, 2014. 5.1.2

[141] Cesar G Victora, Jennifer Harris Requejo, Aluisio J D Barros, Peter Berman, Zulfiqar Bhutta, Ties Boerma, Mickey Chopra, Andres de Francisco, Bernadette Dael-

mans, Elizabeth Hazel, Joy Lawn, Blerta Maliqi, Holly Newby, and Jennifer Bryce. Countdown to 2015: a decade of tracking progress for maternal, newborn, and child survival. *The Lancet*, 387(10032):2049–2059, 2016. ISSN 0140-6736. doi: http://dx.doi.org/10.1016/S0140-6736(15)00519-X. URL http://www.sciencedirect.com/science/article/pii/S014067361500519X. 5.1.2

[142] Eirin Krüger Skaftun, Merima Ali, and Ole Frithjof Norheim. Understanding inequalities in child health in ethiopia: Health achievements are improving in the period 2000–2011. *PLOS ONE*, 9(8):1–10, 08 2014. doi: 10.1371/journal.pone.0106460. URL http://dx.doi.org/10.1371%2Fjournal.pone.0106460. 5.1.2

[143] Atonu Rabbani, Akib Khan, Sifat Yusuf, and Alayne Adams. Trends and determinants of inequities in childhood stunting in bangladesh from 1996/7 to 2014. *International Journal for Equity in Health*, 15(1):186, 2016. ISSN 1475-9276. doi: 10.1186/s12939-016-0477-7. URL http://dx.doi.org/10.1186/s12939-016-0477-7. 5.1.2

[144] World Health Organization. State of inequality: Reproductive, maternal, newborn and child health. Technical report, World Health Organization, 2015. 5.1.2

[145] Marie Ng, K. Ellicott Colson, Nancy Fullman, Laura Dwyer-Lindgren, Tom Achoki, Matthew T. Schneider, Peter Mulenga, Peter Hangoma, Felix Masiye, and Emmanuela Gakidou. Assessing the contribution of malaria vector control and other maternal and child health interventions in reducing all-cause under-five mortality in zambia. *The American Journal of Tropical Medicine and Hygiene*, 2016. doi: 10.4269/ajtmh.15-0315. URL http://www.ajtmh.org/content/early/2016/02/04/ajtmh.15-0315.abstract. 5.1.2

[146] Fernando C Barros, Cesar G Victora, Robert Scherpbier, and Davidson Gwatkin. Socioeconomic inequities in the health and nutrition of children in low/middle income countries. *Rev. Saude Publica*, 44:1–16, 02 2010. 5.1.2

[147] Ahmad Reza Hosseinpoor, Eddy Van Doorslaer, Niko Speybroeck, Mohsen Naghavi, Kazem Mohammad, Reza Majdzadeh, Bahram Delavar, Hamidreza Jamshidi, and Jeanette Vega. Decomposing socioeconomic inequality in infant mortality in iran. *In-*

ternational Journal of Epidemiology, 35(5):1211–1219, 2006. doi: 10.1093/ije/dyl164. URL http://ije.oxfordjournals.org/content/35/5/1211.abstract. 5.1.2

[148] Vu Kien, Hwa-Young Lee, You-Seon Nam, Juhwan Oh, Kim Giang, and Hoang Minh. Trends in socioeconomic inequalities in child malnutrition in vietnam: findings from the multiple indicator cluster surveys, 2000?2011. *Global Health Action*, 9(0), 2016. ISSN 1654-9880. URL http://www.globalhealthaction.net/index.php/gha/article/view/29263. 5.1.2

[149] Owen O'Donnell, Ángel López Nicolás, and Eddy Van Doorslaer. Growing richer and taller: Explaining change in the distribution of child nutritional status during vietnam's economic boom. *Journal of Development Economics*, 88(1):45 – 58, 2009. ISSN 0304-3878. doi: http://dx.doi.org/10.1016/j.jdeveco.2008.01.004. URL http://www.sciencedirect.com/science/article/pii/S0304387808000084. 5.1.2

[150] Florence Jaumotte, Subir Lall, and Chris Papageorgiou. Rising income inequality: Technology, or trade and financial globalization? *IMF Economic Review*, 61(2):271–309, 2013. ISSN 2041-417X. doi: 10.1057/imfer.2013.7. URL http://dx.doi.org/10.1057/imfer.2013.7. 5.1.2

[151] James J. Heckman. Sample selection bias as a specification error. *Econometrica*, 47(1): 153–161, 1979. ISSN 00129682, 14680262. URL http://www.jstor.org/stable/1912352. 1

[152] James Lepkowski. Non-observation error in household surveys in developing countries. In *Household Sample Surveys in Developing and Transition Countries*, pages 149–70. United Nations, 2005. URL http://mdgs.un.org/unsd/hhsurveys/pdf/Chapter_8.pdf. 5.1.3

[153] United Nations. Household Sample Surveys in Developing and Transition Countries. Technical Report Series F No. 96, 2005. 5.1.3

[154] Daniel Kasprzyk. Measurement error in household surveys sources and measurement. Mathematica policy research reports, Mathematica Policy Research, 2005. URL http://EconPapers.repec.org/RePEc:mpr:mprres:d7a25d262708428ba7a6236903ef5b0a. 3

[155] SD Keller, MS Bayliss, JE Ware, MA Hsu, AM Damiano, and TF Goss. Comparison of responses to sf-36 health survey questions with one-week and four-week recall periods. *Health services research*, 32(3):367—384, August 1997. ISSN 0017-9124. URL http://europepmc.org/articles/PMC1070196. 5.1.3

[156] Dennis J. Aigner. Regression with a binary independent variable subject to errors of observation. *Journal of Econometrics*, 1(1):49–59, March 1973. URL https://ideas.repec.org/a/eee/econom/v1y1973i1p49-59.html. 5.1.3

[157] Jeffrey M Woodridge. *Econometric analysis of cross section and panel data*. MIT press, second edition, 2010. page 78. 5.1.3

[158] Kathleen Beegle, Joachim De Weerdt, Jed Friedman, and John Gibson. Methods of household consumption measurement through surveys: Experimental results from Tanzania. *Journal of Development Economics*, 98(1):3–18, 2012. ISSN 0304-3878. doi: http://dx.doi.org/10.1016/j.jdeveco.2011.11.001. URL http://www.sciencedirect.com/science/article/pii/S0304387811001039. Symposium on Measurement and Survey Design. 5.1.3

[159] Menno Pradhan. Welfare analysis with a proxy consumption measure: Evidence from a repeated experiment in Indonesia. *Fiscal Studies*, 30(3–4):391–417, 2009. ISSN 1475-5890. doi: 10.1111/j.1475-5890.2009.00101.x. URL http://dx.doi.org/10.1111/j.1475-5890.2009.00101.x. 5.1.3

[160] D Jolliffe. Measuring absolute and relative poverty: the sensitivity of estimated household consumption to survey design. *Journal of Economic and Social Measurement*, 27 (1), 2001. 5.1.3

[161] Angus Deaton and Valerie Kozel. Data and dogma: The great indian poverty debate. *The World Bank Research Observer*, 20(2):177–199, 2005. doi: 10.1093/wbro/lki009. URL http://wbro.oxfordjournals.org/content/20/2/177.abstract. 5.1.3

[162] Ted Jaditz. Seasonality: economic data and model estimation. *Monthly Labor Review*, 117(12):17–22, 1994. ISSN 00981818, 19374658. URL http://www.jstor.org/stable/41844227. 5.1.3

Attachments: Scientific Papers

PAPER II

Does Free Public Health Care Increase Utilization and Reduce Spending? Heterogeneity and Long Term Effects

Authors: Peter Hangoma, Bjarne Robberstad, and Arild Aakvik

Under Review

II

Does Free Public Health Care Increase Utilization and Reduce Spending? Heterogeneity and Long Term Effects

Peter Hangoma[a,c], Bjarne Robberstad[a], and Arild Aakvik[b]

[a] *Center for International Health, University of Bergen*
[b] *Department of Economics, University of Bergen*
[c] *Unit of Health Economics, School of Public Health, University of Zambia*

Abstract

Zambia removed user fees in public health facilities in 54 out of 72 districts in 2006. The removal was extended to rural areas of previously unaffected districts in 2007. The natural experiment provided by the step-wise implementation of the removal policy as well as five waves of nationally representative household survey data for the period 1998–2010 enables us to investigate short and long term effects of the removal policy on health care utilization and household medical spending. The question is whether free health care increased overall utilization (uptake effect) or merely led to a shift in care from private to public facilities (switching effect). We also examine socioeconomic heterogeneity in the effect of the removal policy on health care utilization.

Using difference-in-difference models, we find that the first wave of user fee removals increased utilization of public health facilities in the short term by 10 percentage points (pp), of which 6.2pp was a result of the increase in overall use of health services (uptake effect) and 3.4pp was due to switching from private facilities. The effect remained strong at 11.1pp in the long term (8.0pp uptake and 3.2pp switching). Overall use of health services (uptake) increased significantly more for individuals whose heads of households were unemployed or had no or less education. Switching on the other hand was driven by individuals who were either formally employed or engaged in farming. The second wave of removals increased utilization by 13.7pp (11.1pp uptake and 2.7pp switching). The probability of incurring any spending reduced significantly for both waves and remained significant in the long term. However, this effect was dampened by an upward pressure on conditional medical spending (spending for those with positive amounts). As a result, the removal policy did not significantly affect total (unconditional) medical spending.

Keywords: User fees; utilization; medical spending; long term; heterogeneity; Zambia

1 Introduction

Low income countries have continued to face substantial health care financing challenges (World Health Organization, 2010; James et al., 2006). This is worsened by often large informal sectors which limit their capacity to implement workable social health insurance schemes (Wagstaff, 2010) or generate sufficient revenue from taxes to finance their health systems (Bitran, 2014). Needless to say, limited financing has resulted in inadequate services on one hand and poor quality services characterized by drug shortages, understaffed health facilities and low staff motivation on the other hand.

Amid these health care financing challenges, low income countries have been urged to remove user fees on grounds that they are significant barriers to health care access for the majority of individuals in these countries who, because of the depth and severity of poverty, experience more health shocks and have limited resources to pay for health services. It is understood that, by increasing medical spending and affecting labor market participation, health shocks have detrimental long term consequences on household income and welfare (Gertler and Gruber, 2002; Islam and Maitra, 2012; Cai et al., 2014; Dhanaraj, 2016). Removing user fees may thus be desirable if it protects household incomes or increases utilization of health services, which in turn may reduce the severity of health shocks or improve health outcomes (Tanaka, 2014; Powell-Jackson et al., 2014).

The push to remove user fees has now gained significant momentum (World Health Organization, 2010; Meessen et al., 2011). So far, this enthusiasm has been based on studies that find that user fees negatively affect utilization (Asfaw et al., 2004; Masiye et al., 2010; Mwabu et al., 1995; Xu et al., 2006; Blas and Limbambala, 2001). This literature parallels studies showing that introduction of user fees, when accompanied with quality improvements may increase utilization (Barber et al., 2004; Litvack and Bodart, 1993; Soucat et al., 1997). However, a number of limitations in both strands of literature have been highlighted (Dzakpasu et al., 2014; Lagarde and Palmer, 2008). Most of these studies rely on administrative data which is generally of poor quality in low income countries. The low quality of administrative data is attributed to both genuine challenges in ensuring data quality and systematic falsification (Sandefur and Glassman, 2015; Lim et al., 2008; Ashraf et al., 2014). Some studies, given the nature of policy changes being investigated, have only been able to conduct before and after studies without a control group–see for example Mwabu et al. (1995); Xu et al. (2006). This makes it difficult to separate the effect of user fees from other concurrent events.

Yet others, e.g., Asfaw et al. (2004), do not have access to user fee policy changes at all. Policy changes may provide plausibly exogenous sources of variation in utilization and enable one to determine the overall impact of a policy and it's heterogeneous effects. As a result of these challenges, even though the removal of user fees is championed as a vehicle to benefit the poor, it still remains unclear whether or not this policy benefits individuals in lower socioeconomic positions more than it does those who are better off.

A few studies that have evaluated some form of user fees using experimental designs yield more internally valid findings (Cohen and Dupas, 2010; Thornton, 2008; Kremer and Miguel, 2007). These studies find lower uptake of interventions in arms with some form of user fees compared to controls. The key limitation of these studies however is that they mainly focus on interventions relating to prevention or vaccination which have lower private valuations, and hence, willingness to pay. It is unclear how utilization would respond in the real world were health seeking behavior is mainly driven by immediate treatment needs, rather than prevention. Recently, a randomized experiment in Ghana where user fees were removed by providing insurance to the intervention arm showed a significant increase in utilization in covered facilities and a drop in medical spending (Powell-Jackson et al., 2014). The challenge however is that experimental studies may lack external validity when one has to imagine large and complex national scale interventions with system wide implications (Acemoglu, 2010). It is important to note that these studies give us partial equilibrium effects, which may not be very informative of the impact of complex nation-wide interventions where general equilibrium effects may dilute or even reverse what would otherwise be significant findings in experimental settings (Acemoglu, 2010). For example, a national scale removal of user fees in public health facilities may lead to lowering of prices associated with private health facilities which may change the expected effect on utilization. Similarly, while individuals may have more substitution possibilities in small studies so that elasticity of an intervention is high, this may not be the case for an intervention implemented at national scale.

An important issue relates to the long term effects of user fee removal. It is unclear whether the widely observed short term effects of user fee removals are sustained in the long term. It is possible that capacity constraints, for example in terms of health workers, may affect quality of health services so that the initial increase weans off in the long term. A study in Zambia and Niger documents that increases in utilization following the removal of user fees were eroded 18 months after the policy (Lagarde

et al., 2012). On the contrary, evidence from Uganda suggests that utilization of health services remained high in the long term (Nabyonga Orem et al., 2013). This finding is consistent with an experimental study from Kenya suggesting that a one off subsidy of health services may sustain utilization in the long term if individuals experience the benefits of such health services (Dupas, 2014).

Similarly, while it is interesting to know whether or not utilization increased in public facilities, it is even more important to know whether this was a result of an increase in overall use of health services (uptake effect) or simply a result of switching from private to public facilities (switching effect), or indeed both. Switching may not generate improvements in health if one assumes that quality of care in public and private facilities is the same. Although it has been argued that private facilities, in their profit maximizing quest, may have pervasive incentives which are potentially harmful to health, there seems to be no apparent differences in quality between private and public facilities (Basu et al., 2012; Powell-Jackson et al., 2015). However, if quality of care is the same but pervasive incentives in private health facilities unnecessarily increase the cost of services, then by eliminating inefficiency in the use of resources, switching may improve social welfare.

Although most of the studies on user fee removal have focused on utilization, perhaps with a view that medical spending falls automatically, there is ample evidence in the health insurance literature suggesting that some form of free care, while increasing utilization, may not reduce health spending (Liu and Zhao, 2014; Fink et al., 2013; Wagstaff et al., 2009; Nguyen, 2012; Ataguba and Goudge, 2012). Some studies actually find that health insurance may even increase financial risk and catastrophic spending (Wagstaff and Lindelow, 2008). It is true, however, that removal of user fees is different from insurance because it eliminates, at least in theory, all health service related spending such as consultation, drug purchases, and examinations while health insurance normally involves some form of cost sharing. In practice, even with the removal of user fees, spending related to health services may still exist, or even increase. This may be due to increases in incidence of informal payments and spending on drugs in the private market. In Uganda, for example, Xu et al. (2006) and Nabyonga Orem et al. (2011) found no evidence of reduced spending following removal of user fees. They reasoned that drug shortages in public health facilities may have forced individuals to rely on the private market which may have pushed medical spending on drugs upwards. The major drawback of these two studies is that they are based on a before and after design without a control group. This makes it difficult to determine the

extent to which the observed changes are attributable to the removal of user fees.

The widespread incidence of informal payments in low income countries is well documented (Lindkvist, 2013; Barber et al., 2004; Falkingham, 2004). Health workers may even reduce effort in order to create a market for such payments (Lindkvist, 2013). Given that user fees may increase health worker motivation (Meessen et al., 2007), one may view informal payments and user fees as substitutes. Evidently, a system of user fees has been shown to minimize informal payments and provide better financial protection in Cambodia (Barber et al., 2004). In Zambia, there is anecdotal evidence of the existence of informal payments after the removal of user fees (Hadley, 2011). A national health expenditure survey conducted in 2013 after the nationwide removal of user fees showed that an appreciable proportion of individuals still incur positive spending (Masiye et al., 2016).

The preceding discussion shows that it is not immediately apparent that removal of user fees reduces medical spending.

The dearth of evidence on the effect of user fees motivated Ridde and Haddad (2009) to conclude that "African public health officials and decision makers are worried about the relationship between abolishing user fees and health care financing, and much remains to be done to provide them with the evidence they require."

By combining several waves of huge nationally representative household survey data in Zambia for the period 1998–2010 and the natural experiment provided by the stepwise implementation of the removal policy, we overcome some of the methodological and data challenges of the literature attempting to study complex user fee policy interventions implemented at national scale. Our identification strategy exploits the fact that in April 2006, the government of Zambia removed user fees in all public health facilities in 54 out of 72 districts classified as rural. The removal of user fees was extended to rural areas of the remaining 18 previously unaffected districts in June 2007 (MoH, 2007). Thus, only urban areas of the 18 districts remained unaffected by both waves of the removal policy. Our empirical approach relies on difference-in-difference (DD) models. Multiple pre-treament periods enable us to assess the validity of our identifying assumption and carry out a number of robustness checks.

This study contributes to the debate on free health care in general, and removal of user fees in particular, in at least five ways. First, our data enables us to examine whether or not the removal of user fees affected overall utilization of health services. Previous studies in Zambia have only examined the effect on utilization of public health facilities and found that there was a substantial increase following the first wave of

removals (Masiye et al., 2010; Lagarde et al., 2012; Onde, 2009). However, it is possible that these increases were merely a result of individuals switching from private to public facilities, as opposed to an increase in the use of health services by individuals who were previously excluded by virtue of being poor. Second, we are able to examine heterogeneous effects of the removal policy by socioeconomic status. While we may not be able to examine heterogeneities by income or total consumption because these variable were potentially impacted by the removal policy, our rich household survey data has other key socioeconomic variables such as occupational status and education which were arguably not affected by the removal policy. In other words, we are able to examine whether or not removal of user fees elicited a higher utilization response from individuals from lower socioeconomic backgrounds as the policy intended. Third, we study the long term effect of the removal policy. This shades light on the ability of the removal policy to sustain gains in utilization and medical spending. Fifth, we provide evidence on the extent to which the removal policy affected the proportion of individuals incurring medical spending, and the amount of spending. This is an important starting point in discussing financial risk protection of the user fee removal policy.

Our result show that the increase in utilization of public health facilities was mainly driven by an increase in the overall use of health services. The first wave of user fee removals increased utilization of public health facilities in the short term by 10 percentage points (pp), of which 3.4pp was due to switching from private health facilities (switching effect) and 6.2pp was a result of the increase in overall use of health services (uptake effect). The effect remained strong at 11.1 pp in the long term, four years later (3.2pp switching effect and 8.0pp uptake effect). Similarly, the second wave of removals increased utilization by 13.7pp (2.7pp switching effect and 11.1pp uptake effect). With regard to medical spending, the proportion of individuals incurring any spending fell in the short term and this effect was maintained in the long term. However, there was an upward pressure on the amount of spending for those individuals still incurring any spending (conditional spending). This means that the policy had a significant effect on the extensive margin but not the intensive margin and the result was that total (unconditional) health spending was unaffected.

The rest of this paper is organized as follows: Section 2 provides a short overview of the setting of this study and the user fee removal policy. Section 3.1 discusses data and identification. The empirical specification is given in Section 4, results in Section 5, and robustness checks in Section 6. We discuss our findings in Section 7 and conclude

in Section 8.

2 Context and User Fee Removal

Zambia is classified as a lower middle income country with GNP per capita of USD 1,810 (World Bank, 2015). It is one of the most unequal countries in terms of income, and social services (UNDP, 2014). In 2010, poverty levels were estimated at 60.5% with more than 83% of the labor force employed in the informal sector (CSO, 2012). The size and extent of informal economic activities possess a challenge to attempts by the government to expand the tax base. The disease burden is also high with malaria, tuberculosis and HIV exerting a severe strain on the health system. The burden of non-communicable diseases is also on the increase (Institute for Health Metrics and Evaluation, Human Development Network, The World Bank, 2013).

The Health care market in Zambia is almost completely dominated by public facilities–government health and faith based (mission) facilities–which cater for more than 90% of the market (CSO, 2012, 2011). Private facilities cater for up to 10% of the market and they are mainly concentrated in urban areas, with little or no presence in rural areas. Faith based health facilities are mostly located in rural areas and are classified as public because government provides most of their health staff and sometimes allocates funding. In addition, government policies such as introduction or removal of user fees affect them equally (MoH, 2007).

From Independence in 1964, Zambia had a tax financed health care system were health services were provided freely. As part of the a number of structural reforms in the early 1990s however, user fees were introduced with exemption for children below five years of age and adults above 65 years. User fees were not uniform but varied across the country and were agreed upon by health facility management and community representatives taking into account the local economy of the health facility catchment area.

The country now has a mixed health care financing system with a heavy reliance on external financing. The National Health Accounts (NHA) figures show that between a third to half of total health expenditure (THE) is from external sources (World Health Organization, 2015; Ministry of Health, 2009). General government expenditure on health (GGEH) as a share of general government expenditure (GGE) averaged 12% between 2002 and 2013.

Zambia was granted debt relief in 2005 under the Highly Indebted Poor Countries

Initiative (HIPC). Savings from debt servicing and it's long held motto of striving to ensure universal access to health services as well as providing financial protection motivated the government to make health services free (MoH, 2007). Surprisingly, the NHA data also shows that out of pocket expenditure (OPE) as a share of THE was significantly higher after 2006 (World Health Organization, 2015), a year marking the first wave of removal of user fees. Specifically, in April 2006, Zambia removed user fees at the primary level in all public health care facilities in 54 districts designated as rural, leaving the rest of the districts (18) that were classified as urban, unaffected. Moreover, individuals who went through the referral system continued to be exempt at higher levels of care. User fees were defined to include fees for registration, consultation, outpatient and inpatient care, X-ray, and laboratory tests. Figure 1 show the districts which were affected and unaffected by the first wave of removal of user fees. While user fees were removed at primary level only (clinics and level 1 hospitals) in public facilities, sweeping exceptions were given for districts to offer free care even at higher levels. For example, services were to be free at higher level hospitals (level 2 and 3) in districts without level 1 hospitals (MoH, 2007).

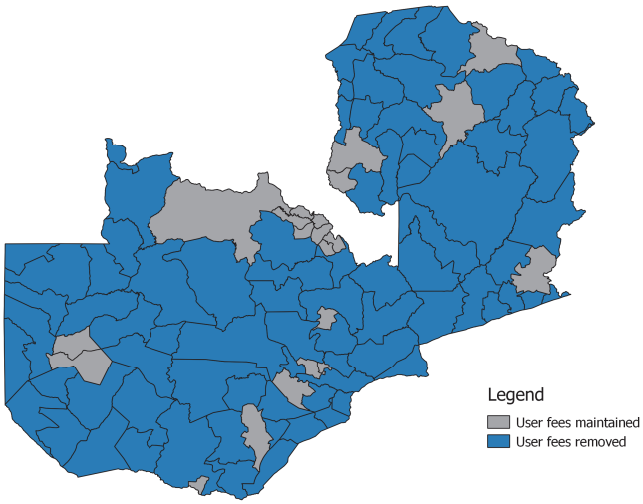


Figure 1: Zambia. Map shows districts were user fees were removed during the first wave of user fee removals. Three of the 18 district were each recently split into two. Hence the number of district were user fees were maintained on this map will count as 21 instead of 18.

The following year, in June 2007, the government acknowledged the shortcoming of

removing user fees based on classifying the whole district as either rural or urban, and that there are areas within urban districts that were rural and vice versa. With this understanding, they extended the removal of user fees to rural areas of the 18 districts that were previously unaffected. In particular, areas where user fees were removed depended on whether or not the district was located along the line on rail (railway). For those along the line of rail, user fees where removed in health facilities located more than 20 km away from the district administration. Similarly, user fees were removed in health facilities located more than 15 km away from the district administration for districts that were not located along the line of rail. Implicit in the removal policy is the idea that rural areas are located in the peripheries of the districts.

In 2012, user fees were finally removed throughout the country in all public health facilities.

3 Data and Identification

3.1 Data Sources, Sample, and Outcome Variables

Our data are from the 1998, 2002, 2004, 2006 and 2010 Living Conditions Monitoring Survey (LCMS). The LCMS is a large nationally representative repeated cross-sectional household survey conducted by the Central Statistical Office of Zambia (CSO). Although LCMS survey were also conducted before 1998, it was not possible to use these data because of changes in district boundaries making it impossible to isolate treated from control districts.

For medical spending, we do not include the 1998 data due to differences in components of medical spending that were collected in 1998 compared to latter surveys. While the 1998 survey asked about spending on a number of health services in the two weeks prior to the survey, the 2002 and latter surveys only asked about total spending in the two weeks prior to the survey.[1] It has been shown that more dissagregation and longer list of items, as is the case in the 1998 survey, yield higher levels of expenditures for similar households compared to less dissagregation (Beegle et al., 2012; Pradhan, 2009; Jolliffe, 2001).

Our analysis focuses on individuals who reported any sickness or injury in the two weeks prior to the survey. We restrict our sample to individuals aged between 5 and 65. This is because individuals under 5 years and those over 65 were exempt from

[1]This question existed in the 1998 survey but only referred to expenditure on self medication

user fees even before the removal. The percentage (number) of individuals between the age of 5 and 65 reporting sickness or injury was 9% (7,018) in 1998, 12% (6,482) in 2002, 8% (8,182) in 2004, 7% (6,372) in 2006, and 11% (10,003) in 2010. Individuals who reported any sickness or injury were asked whether or not they consulted any health facility as a result of the sickness or injury. Those who reported consulting were then asked about the type of facility they consulted. Using this information, we define three measures of utilization. The first one, is an indicator variable equal to one if an individual reported consulting a public facility and zero if not. Individuals are said to have consulted a public facility if they visited a government health post, clinic, hospital or a faith based health facility for the illness or injury.[2] This measures total response of utilization of public facilities and it may capture both the fact that there are more or less people using health services irrespective of provider (uptake effect) and that their is substitution between public and private facilities (switching effect). Thus, our second measure of utilization is also an indicator variable equal to one if an individual reported consulting a private health care facility and zero otherwise. This captures the switching effect. The third measure is equal to one if an individual reported having made any consultation (public or private) and it measures the overall increase or reduction in health care use following the removal of user fees.

Medical spending in our data refers to the amount spent on consultation, medical examinations, drugs, and any form of self medication in the past two weeks for individuals who reported being sick or injured. We define two measures of spending, one that looks at the proportion of individuals that incurred any spending (extensive margin) and the other measure which assesses the level of spending conditional on it being positive (intensive margin).

All outcomes are conditional on reporting any illness or injury in the two weeks prior to the survey.

3.2 Identification and Data Description

We define the first wave of removal of user fees in early 2006 as Treatment 1 (T1) and the second wave of removals in June 2007 as Treatment 2 (T2). These two waves of removals partitioned the whole country such that only urban areas of the 18 districts remained unaffected, and these areas represent our control group. Individuals who reside in the areas were T1 was effected are defined as the T1 group. Using the T1

[2]See Section 2 for a discussion of why faith based health facilities are classified as public health facilities

group and the control, we are able to identify the short term (2006) and long term (2010) effects of the first wave of removals, T1. Individuals residing in the rural areas of the 18 urban districts who were affected by the second wave of removals, T2, in 2007 are defined as the T2 group. This treatment group enables us to identify the effect, in 2010, of the second wave of removals. Table 1 briefly describes our treatment and control groups.

Table 1: Description of Treatment and Control groups

Group	**Description**	**N**
T1	Individuals residing in the 54 districts were user fees were removed in April 2006 at the primary level in all public health care facilities. User fees were maintained in the rest of the districts–18 districts.	23,403
T2	Individuals residing in the rural areas of the 18 districts. They were affected by the second wave of removals in June 2007.[a]	4,140
Control	Individuals residing in urban areas of the 18 districts. These were not affected by the two waves of the removal policy in 2006 and 2007.	10,514

[a] The precise identification of these areas would require GPS coordinates which were unfortunately not collected in the surveys. We use the rural (urban) definition of the Central Statistical Office (CSO) to classify the areas in the 18 districts as treated (control). Our definition appears to distinguish treated and untreated areas in the 18 district very well. See Appendix A.1 for details.

Given multiple waves of data, we exploit geographical variation in the removal policy and use difference-in-difference models to estimate the short term (2006) and long term (2010) effects of providing free care in public health facilities. Since we are not able to observe those who actually received free care, this is an intention to treat (ITT) effect of the removal policy.[3]

Our identifying assumption is that, in the absence of the removal policy, potential outcomes in areas where user fees were removed would have followed the same trend as non-treatment outcomes in the areas where user fees were not removed. A departure from trend is counted as the effect of the policy. This identifying assumption

[3]Some individuals reported having paid for care in primary public health care facilities even after the national wide removal of user fees in 2013 (Masiye et al., 2016). It is not clear if these were informal charges or other charges introduced by health facilities following removal of user fees. Hadley (2011) documents cases were some health facilities charged different types of fees after the official fees were removed. ITT would only equal the average treatment on the treated (ATT) if there was full compliance. In a world were compliance is a problem, for example due to informal fees, ITT is a more policy relevant effect.

is fundamental in the difference-in-difference framework and it is called the *parallel trends* assumption. *Parallel trends* does not require that treated and control groups are the same, but that pre-treatment outcomes follow similar trends (Godlonton and Okeke, 2016). Formally, this assumption is assessed by checking that the differences in outcomes between the treated and control groups are the same at all time periods before treatment (pre-treatment period). This implies that a difference of the differences in outcomes between any two time points in the pre-treatment period should be zero. In other words, in a difference-in-difference model, if the pre-treatment interaction (treatment) effects are included, they should be zero, or statistically insignificant. We conduct placebo tests by including pre-treatment treatment effects in our models. Before then, we can visually assess the plausibility of this assumption.

3.2.1 Assessment of Pre-treatment Trends

We present trends in average levels of utilization and medical spending for the T1 and control group in Figure 2. The associated confidence intervals are also shown. The plots on the left of the figure show the percentage of individuals utilizing public health facilities (top) and private health facilities (bottom). On the right of the figure, we show the percentage of individuals incurring any medical spending (top) and conditional spending in natural logarithms (bottom).

Utilization of both public and private health facilities in the two groups, T1 and control, exhibited similar trends throughout the pre-treatment period (1998–2004). In the first part of the pre-treatment period, 1998–2002, utilization of public facilities significantly increased in both the T1 and control group while that of private facilities remained almost unchanged in both groups. Similarly, with consideration of confidence intervals, utilization levels of both public and private facilities did not change in any significant way for both groups in 2004.

Regarding medical spending, both the percentage of individuals incurring any spending and log of medical spending in the two groups appears to follow similar trends during the pre-treatment period, 2002–2004.[4]

Thus, outcomes in the T1 and control groups appear to follow similar pre-treatment trends.

[4]Note that, as discussed in Section 3.1, we only use two pre-treatment data periods, 2002 and 2004, for spending.

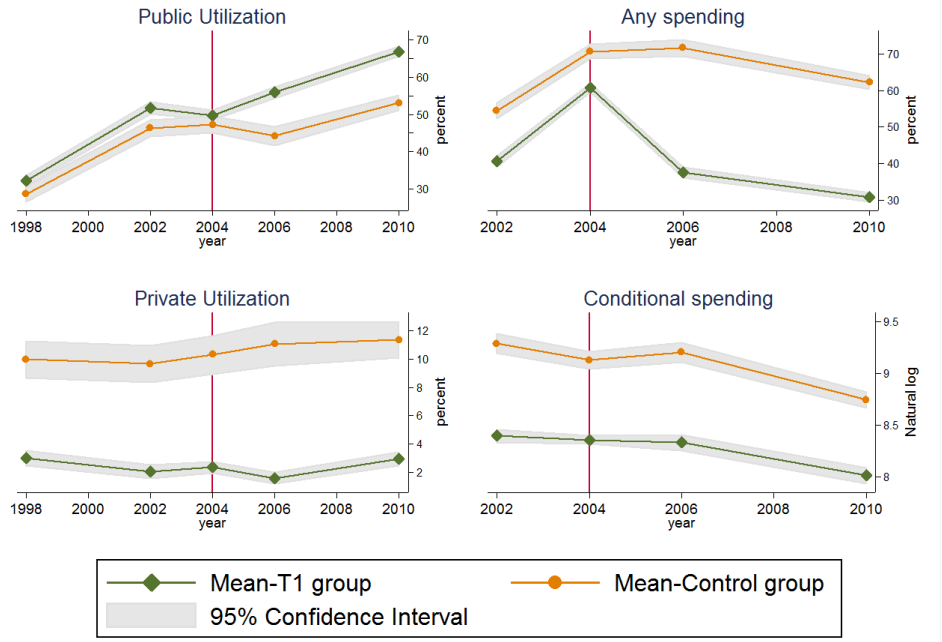


Figure 2: Figure shows trends in average utilization and spending, and the associated 95% confidence intervals, for T1 group and control

Turning to the T2 group, the overlapping confidence intervals for utilization of public facilities in Figure 3 throughout the pre-treatment period (1998–2006) suggests that utilization rates of public facilities in the T2 group and control were the same. This may not be surprising given that some individuals in the T2 group and those in the control belong to the same district, only that those in the T2 group are located further from the district administration.[5] The fact that health services and other social amenities are controlled by the same district managers makes the possibility of similar trends plausible. We note however that, although not significant, utilization of public facilities in the T2 group went slightly above that of the control in 2006. It is worth mentioning that there is a possibility spillover were some individuals in the T2 group, by virtue of living further from their district administration–as specified by the removal policy–may have had their closest health facilities in the T1 areas, which were treated in 2006. The figure however shows that this possible spillover effect is not

[5]Recall that for each of the 18 districts, the second wave of removals, T2, unlike T1, removed user fees in only certain parts of each of the 18 districts. These parts were considered sufficiently rural by virtue of being far from the district administration.

significant to invalidate the parallel trends assumption. Utilization of private facilities also appears to follow similar trends in the two groups in the pre-treatment periods.

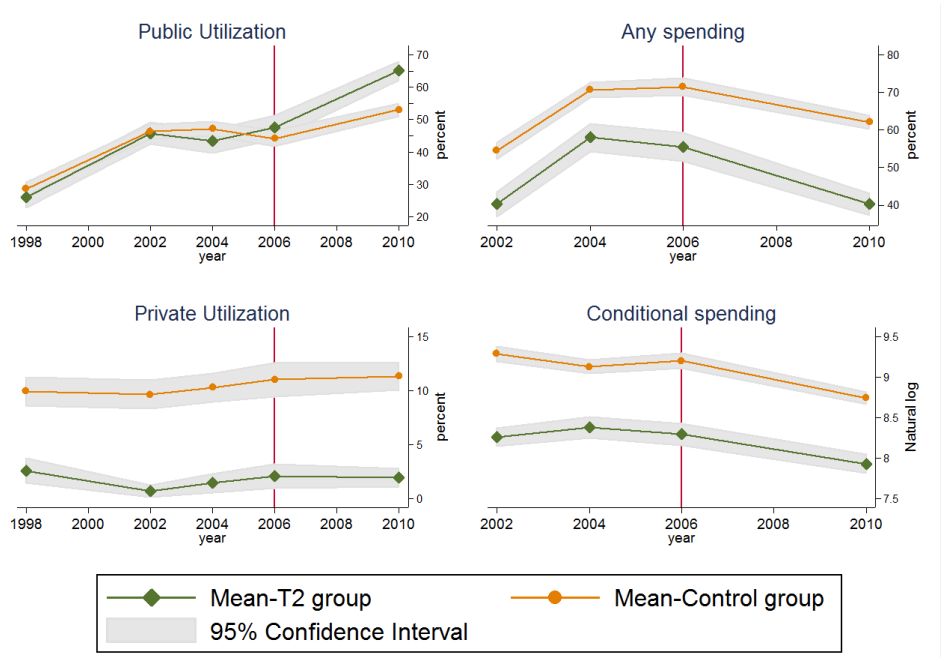


Figure 3: Figure shows trends in average utilization and spending, and the associated 95% confidence intervals, for T2 group and control

3.2.2 Baseline Covariates and Changes Over Time

The T1 and control group are generally similar in terms of household and demographic characteristics such as household size, age, sex, and marital status but differ in characteristics such as educational level, occupational status and distance to a health facility (Table 2, Column (1) and (2)). For instance, 12.2% of household heads had no education in the T1 group compared to 4% in the control group. Only 16.9% of individuals in the T1 group had their nearest health facility with 1km compared to 45.6% in the control group. Although the T1 group had significantly fewer household heads who were unemployed, it also had a substantially lower proportion of household heads formally employed. Differences in demographic and socioeconomic characteristics between treated and control may explain differences in utilization and medical spending. However, these differences do not endanger the plausibility of the parallel trends assumption.

The parallel trends assumption is more likely to hold if demographic and socioeconomic characteristics (covariates) in treated and control groups change in the same way overtime, especially in the post-treatment period where non-parallel trends in outcomes emerge. In other words, if covariates in the two groups change differently in the post-treatment period, they could actually be the ones driving differences in utilization and medical spending that we may attribute to the user fee removal policy. Column (3)–(6) tests the hypothesis that characteristics in the T1 and control group changed in the same way between baseline (2004) and each survey year. Since we have seen from the previous section that trends in outcome were parallel up to the last pre-treatment year (2004) and difference emerged in 2006 and 2010, it is interesting to see whether characteristics in the T1 and control were changing differently between 2004 and 2010. Column (5) and (6) shows that almost all characteristics did not change differently.

Table 2: Baseline and changes in characteristics: T1 Group vs Control

	Baseline (2004) Average		Change: 1998-2004	Change: 2002-2004	Change: 2004-2006	Change: 2004-2010
	T1 Group	Control	$\Delta T1 = \Delta Control$ (p-value)	$\Delta T1 = \Delta Control$ (p-value)	$\Delta T1 = \Delta Control$ (p-value)	$\Delta T1 = \Delta Control$ (p-value)
Variable	(1)	(2)	(3)	(4)	(5)	(6)
Male	0.428	0.453	0.262	0.876	0.796	0.153
HH Male	0.784	0.750	0.045	0.348	0.424	0.779
H Size	6.165	6.072	0.786	0.670	0.882	0.161
Age	27.382	27.397	0.120	0.418	0.857	0.129
HH Married	0.768	0.659	0.013	0.146	0.100	0.217
HH College	0.048	0.121	0.866	0.725	0.978	0.588
HH Secondary	0.287	0.562	0.157	0.130	0.575	0.029
HH Primary	0.542	0.277	0.052	0.446	0.750	0.383
HH No Education	0.122	0.040	0.359	0.075	0.544	0.104
HH Employed	0.143	0.418	0.069	0.004	0.773	0.165
HH Self Employed	0.090	0.306	0.054	0.055	0.394	0.450
HH Farming	0.707	0.071	0.831	0.000	0.143	0.006
HH Not Working	0.061	0.205	0.479	0.934	0.292	0.189
Distance to Nearest Health Facility:						
$\leq$ 1 Km	0.169	0.456	0.007	0.381	0.616	0.839
1-10 Kms	0.612	0.529	0.029	0.159	0.462	0.172
11-20 Kms	0.146	0.000	0.980	0.241	0.182	0.003
$\geq$ 21 Kms	0.073	0.015	0.326	0.605	0.807	0.046

Note: The table shows characteristics of individuals aged 5 to 65 who reported being sick or injured in the two weeks prior to each survey in the T1 group and control. The first and second columns report means of covariates at baseline (2004) for treated and control group, respectively. The rest of the columns report p-values that compare changes in the control and treated group from baseline to the relevant survey year. The p-values where computed using a simple difference on difference model of each characteristic (covariate). HH=Household Head and H=Household. All variables are dummies except H Size, and Age.

Table 3 presents a comparison of characteristics in the T2 and control group. The picture is generally similar to that of the T1 and control group; although the T2 and control groups are broadly the same in terms of household and demographic

characteristics such as household size, age, sex, and marital status, the T2 group is worse of in terms of characteristics such as educational level, occupational status and distance to a health facility (Table 3, Column (1) and (2)). However, these characteristics generally change in the same way in the two groups over time.

To ensure that any differential changes in characteristics do not drive outcomes, we control for them in a difference in difference regression models. Given that these characteristics were changing in the same way, we anticipate that there would not be significant difference in results between models with and without covariates.

Table 3: Baseline and changes in characteristics: T2 Group vs Control

	Baseline (2004) Average		Change: 1998-2004	Change: 2002-2004	Change: 2004-2006	Change: 2004-2010
	T2 Group	Control	$\Delta T2 = \Delta Control$ (p-value)	$\Delta T2 = \Delta Control$ (p-value)	$\Delta T2 = \Delta Control$ (p-value)	$\Delta T2 = \Delta Control$ (p-value)
Variable	(1)	(2)	(3)	(4)	(5)	(6)
Male	0.399	0.453	0.427	0.357	0.061	0.023
HH Male	0.811	0.750	0.062	0.116	0.009	0.235
H Size	5.749	6.072	0.890	0.483	0.897	0.101
Age	25.299	27.397	0.229	0.595	0.859	0.313
HH Married	0.794	0.659	0.016	0.083	0.024	0.260
HH College	0.016	0.121	0.540	0.930	0.797	0.416
HH Secondary	0.636	0.277	0.648	0.413	0.981	0.621
HH Primary	0.253	0.562	0.888	0.144	0.534	0.358
HH No Education	0.095	0.040	0.794	0.730	0.439	0.274
HH Employed	0.131	0.418	0.087	0.114	0.900	0.893
HH Self Employed	0.123	0.306	0.344	0.919	0.168	0.819
HH Farming	0.684	0.071	0.365	0.073	0.520	0.866
HH Not Working	0.061	0.205	0.588	0.929	0.680	0.655
Distance to Nearest Health Facility:						
$\leq$ 1 Km	0.178	0.456	0.063	0.788	0.020	0.691
1-10 Kms	0.644	0.529	0.029	0.159	0.111	0.342
11-20 Kms	0.131	0.000	0.274	0.513	0.846	0.253
$\geq$ 21 Kms	0.047	0.015	0.711	0.501	0.704	0.183

Note: The table shows characteristics of individuals aged 5 to 65 who reported being sick or injured in the two weeks prior to each survey in the T2 group and control. The first and second columns report means of covariates at baseline (2004) for treated and control group, respectively. The rest of the columns report p-values that compare changes in the control and treated group from baseline to the relevant survey year. The p-values where computed using a simple difference on difference model of each characteristic (covariate). HH=Household Head and H=Household. All variables are dummies except H Size, and Age.

4 Empirical Specification

Our econometric model for evaluating the impact of the two waves of the removal of user fees, T1 and T2, is a difference-in-difference model of the form:

$$\begin{aligned} y_{it} &= \alpha_0 + \alpha_1 \text{yr1998} + \alpha_2 \text{yr2002} + \alpha_3 \text{yr2006} + \alpha_4 \text{yr2010} + \delta_1 \text{T1}_i + \delta_2 \text{T2}_i \\ &+ \delta_3(\text{T1}_i \times \text{yr2006}) + \delta_4(\text{T1}_i \times \text{yr2010}) + \delta_5(\text{T2}_i \times \text{yr2010}) \\ &+ \delta_6(\text{T1}_i \times \text{yr1998}) + \delta_7(\text{T1}_i \times \text{yr2002}) \\ &+ \delta_8(\text{T2}_i \times \text{yr1998}) + \delta_9(\text{T2}_i \times \text{yr2002}) + \delta_{10}(\text{T2}_i \times \text{yr2006}) \\ &+ \boldsymbol{\beta}' X_{it} + \boldsymbol{\gamma}_1 \sum_{r}^{R-1} \text{P}_r + \gamma_2 \sum_{r}^{R-1} \sum_{t=1998,2002}^{2006,2010} (\text{P}_r \times \text{yr}_t) + \epsilon_{ijt} \end{aligned} \tag{1}$$

where y_{it} is the observed outcome for individual i in year t. The variables yr2002, yr2006 and yr2010 are year dummies. They account for changes that may have affected utilization and medical spending in both treatment and control groups, typically national level policies or economic shocks. The baseline year is 2004. T1_i is a treatment indicator equal to 1 if individual i is in the T1 group, i.e., resides in any of the 54 treated districts. Similarly, T2_i is 1 if they reside in the rural areas of the other 18 districts. T1_i and T2_i captures the influence of unobserved factors which are specific to these areas and have the potential to explain differences in utilization and medical spending between each treatment group and the control. X_{it} is a set of individual and household observable characteristics of i, e.g education, employment status, household characteristics, distance to health facility, etc. They account for any time varying individual and household characteristics that may have had differential impact on outcomes of treated and control groups. P_r is the rth province dummy, for $R = 9$, while $\text{P}_r \times \text{yr}_t$ are province by year dummies. They capture regional level fixed effects and time varying shocks that may have affected the treated and control group differently. Supposing this is a linear model, the short term difference-in-difference effect of T1 is given by the coefficient of the first interaction term δ_3, while the coefficient on second interaction term, δ_4, gives the long term effect of T1. We measure the effect of the second wave of user fee removals, T2, by the coefficient, δ_5. As a placebo test to formally assess the plausibility of the parallel trends assumption, a full set of pre-treatment interaction effects for both T1 and T2 are included. δ_6 and δ_7 should be zero if the T1 group and control followed similar pre-treatment trends. In the same vein, δ_8, δ_9, and δ_{10} should not be statistically different from zero if the T2 group and control followed similar pre-treatment trends.

To ensure that standard errors are not underestimated since the policy was implemented at the district level while the data is at the individual level, we clustered the standard errors at the district level (Bertrand et al., 2004).

Utilization

All three measures of utilization (public, private and any utilization) are binary. For each of these outcomes, we fitted a Linear Probability Model (LPM) to Equation (1). Thus, for each of the three outcomes, the effect of two waves of the removal policy are simply the coefficients, δ_3, δ_4, and δ_5. To save space, we only report these effects and the pre-treatment interaction coefficients δ_6, δ_7, δ_8, δ_9, and δ_{10}.

To understand the distributional impact of the removal policy, we examine how utilization of various individuals across the socioeconomic distribution was affected. These socioeconomic variables should not have been affected by the removal policy. Thus, rather than focusing on income or household consumption, we look at educational level and occupation of the head of household. We categorize these variables and estimate a difference-in-difference model for each of these categories (sub-samples). Focusing on T1, we estimate the overall effect for each sub-sample, as opposed to short and long term effects as we did in Equation (1), in order to concentrate on examining heterogeneities. Thus, for each of the sub-samples of these variables, we fit a difference-in-difference model of the form:

$$\begin{aligned} y_{it} &= \theta_0 + \theta_1 \mathrm{T1}_i + \theta_2 \mathrm{post}_i + \theta_3(\mathrm{T1}_i \times \mathrm{post}_i) + \theta_4 \mathrm{yr1998} + \theta_5 \mathrm{yr2002} \\ &+ \boldsymbol{\beta}' X_{it} + \boldsymbol{\gamma}_1 \sum_{r}^{R-1} \mathrm{P}_r + \gamma_2 \sum_{r}^{R-1} \sum_{t=1998,2002}^{2006,2010} (\mathrm{P}_r \times \mathrm{yr}_t) + \epsilon_{ijt} \end{aligned} \tag{2}$$

where post_i is 1 if an individual was observed in 2006 or 2010. Our parameter of interest, which we report for each sub-sample, is θ_3. This analysis will give us an insight of which socioeconomic group began utilizing any health services after the removal (uptake effect) and which ones could have been switching from private to public facilities where user fees were removed. Thus, we focused on two outcomes, any utilization and private utilization.

Medical Spending

As is common in household expenditure data, our medical spending data has many zeros and is particularly skewed. This is especially true because the removal policy eliminated medical spending for some individuals. Although OLS on log transformed spending is simple and deals with the skewness problem, it yields biased estimates because it ignores the existence of zero expenditure. It is important to study the impact of the policy on the full distribution of medical spending. To achieve this, we examined the policy effect on two margins of the expenditure distribution, the extensive margin–whether or not an individual incurred any spending–and the intensive margin–the amount of spending conditional on it being positive. We then combined these two effects to yield a measure of the impact of the policy on the whole (unconditional) distribution of medical spending. This was achieved using a two-part model (TPM).

In the TPM, the extensive margin (first part) is modeled by fitting a probit model to Equation (1) and the intensive margin (second part) similarly fits Equation (1) using generalized linear models (GLM). The Box Cox test is used to select the link function while the family is chosen using the modified Park test. This lead us to a gamma family with a logarithmic link function.

Because probit is a non-linear model, the interaction or marginal effect of the policy are not given by the interaction terms in Equation (1) (Ai and Norton, 2003). For the extensive margin (first part) thus, each interaction effect, e.g for T1 in 2006, was computed as the following double difference:

$$\begin{aligned} \frac{\Delta^2 E}{\Delta yr2006 \Delta T1} = & \quad [\Phi\{\text{yr2006} = 1, \text{T1} = 1, X\} - \Phi\{\text{yr2006} = 0, \text{T1} = 1, X\}] \\ - & \quad [\Phi\{\text{yr2006} = 1, \text{T1} = 0, X\} - \Phi\{\text{yr2006} = 0, \text{T1} = 0, X\}] \end{aligned} \tag{3}$$

where Φ is the normal cumulative distribution function and the matrix X contains all covariates including region, other interaction terms and year dummies.

For the second part (the GLM), the interaction effect was similarly computed as in Equation (3), although Φ is simply an exponent.

The overall effect on medical spending (unconditional medical spending) was estimated by combining the effect of the removal policy on each of two parts of the TPM. To see this, note that the unconditional expected value of medical spending is the joint expected value of medical spending, $E[y]$, which in turn is given by the product of the probability of incurring any spending-the probit part-and the average spending, given that spending is positive-the GLM part:

$$E[y] = \Phi(\cdot) \times E[y|y > 0] \tag{4}$$

The impact of the policy is given by how it changes this joint expeted value. For example, the effect of T1 on overall medical spending in 2006 is given by taking the double difference (similar to Equation (3)) on Equation (1). Taking the double difference is a binary variable equivalent of differentiating the joint expectation twice.[6]

Standard errors a computed using a bootstrap procedure with 1000 replications that accounts for clustering at the district level.

5 Results

5.1 Impact on Individual Utilization

Overall Impact

Results indicate that the removal of user fees in public health facilities increased the overall utilization of health services (uptake effect), in addition to shifting use from private to public facilities (switching effect). However, the uptake effect was much greater, and stronger, than the switching effect (Table 4). The first wave of user fee removals, T1, increased overall utilization (uptake) of health services by 6.2pp in the short term (Column (1)). The effect strengthened to 8pp in the long term. However, utilization of public facilities increased more than the increase in overall utilization of health services in both the short and long term. Column (2) shows that utilization of public facilities increased by 10pp in the short term and the effect was sustained at 11.1pp in the long term. The extra increase in the utilization of public facilities was a result of switching from private to public facilities (Column(3)). Utilization of private facilities in the T1 group reduced by 3.4pp in the short term with the effect sustained at 3.2pp in the long term.

The second wave of removals (T2) increased the overall utilization of health services by 11.1pp. Utilization of public facilities increased by 13.7pp of which 2.6pp was due to switching from private facilities, although the switching effect was not significant. This insignificance is expected given the small proportion of individuals utilizing private facilities in T2 areas which renders standard errors to be imprecisely estimated.

[6] See Frondel and Vance (2013) for a discussion on interaction effects in non-linear two-part models

Table 4: Short and long term effect of user fee removal on overall, public, and private facility utilization

	Overall Utilization (1)	**Public Utilization** (2)	**Private Utilization** (3)
Short term effect			
T1 × yr2006	0.062(0.029)**	0.100(0.030)***	−0.034(0.017)*
Long term effects			
T1 × yr2010	0.080(0.032)**	0.111(0.035)***	−0.032(0.014)**
T2 × yr2010	0.111(0.048)**	0.137(0.047)***	−0.027(0.016)
Pre-treatment effects			
T1 × yr1998	0.000(0.036)	0.004(0.033)	−0.004(0.022)
T2 × yr1998	0.018(0.064)	0.006(0.064)	0.011(0.023)
T1 × yr2002	−0.004(0.030)	0.009(0.032)	−0.013(0.016)
T2 × yr2002	−0.003(0.061)	0.014(0.059)	−0.011(0.016)
T2 × yr2006	0.078(0.065)	0.093(0.061)	−0.015(0.023)
Observations	31,887	31,887	31,887
T1 group baseline mean	0.521	0.497	0.024
T2 group baseline mean	0.447	0.433	0.014
Control group baseline mean	0.575	0.472	0.103

Note: Table shows the effect of the first and second wave of the removal policy, T1 and T2 respectively, on utilization of all facilities (Column 1), public facilities (Column 2) and private facilities (Column 3). Estimates are from DD linear probability models (LPM) corresponding to Equation (1) and are interpreted as percentage points. Standard errors clustered at the district level are reported in parentheses. The lower panel gives the baseline (2004) mean of each outcome for treated and control groups. All models include a full set of year dummies, regional effects, and region by year interaction effects. Covariates include distance to health facility, household size, sex, age, and age squared of an individual, as well as household head characteristics such as occupational status, marital status, and educational level.
* Significant at 10%; ** Significant at 5%; *** Significant at 1%

All pre-treatment "treatment effects" for the T1 are zero lending support to the descriptive evidence of parallel trends in Section 3.2.1. The effect of the first wave of removals, T1, is thus identified. Similarly, all pre-treatment effects for T2 are not significantly different from zero. This suggests that utilization in the T2 group and the control also followed approximately parallel trends. However, though not statistically different from zero, the size of pre-treatment effect for T2 in 2006 is large suggesting, as discussed in Section 3.2.1, that there may have been spillover where individuals in the T2 group could have accessed free services in the T1 areas in 2006. By definition, individuals from the T2 group are located far from their district centers (MoH, 2007), and hence closer to the T1 areas, which happened to have had user fees removed in 2006. Note however that the descriptive in section 3.2.1, and standard errors in Table 4, indicate that this possible spillover effect was not significant enough to invalidate

the parallel trends assumption. Results for the second wave of removal, T2, should, nonetheless, be interpreted with caution.

Heterogeneous Impact

Table 5 shows how individuals from different socioeconomic backgrounds responded to the removal of user fees. Estimates are presented for sub-samples of education level and occupation status of the head of household. Each estimate is derived from running Equation 2 on each sub-sample. To save space, we only report θ_3. The removal of user fees in public health facilities increased overall utilization of health services more for individuals from lower socioeconomic background than their better off counterparts (Table 5). For individuals from higher socioeconomic background, the policy led to a shift in care seeking from private to public facilities.

Beginning with education level of head of household, the increase in overall use of health facilities was greater the lower the educational level of the head of household (Column (1)). Specifically, individuals whose heads of households had no education increase their utilization most (10.6pp). Overall use of health services significantly increased by 6.2pp for individuals whose household heads only had primary education and, although positive, the increase was not significant for individuals whose household heads had a secondary or college education. On the other hand, the switching effect was generally driven by individuals coming from household heads with higher education (Column (2)).

For occupation status, overall use of health services for individuals whose household heads were not employed significantly increased by 11.2pp. Although positive, the overall use of health services did not increase significantly for individuals whose heads of households were formally or self employed. For individual engaged in farming, the policy did not affect overall utilization of health services. It however shifted their care seeking from private to public facilities. Individuals whose household heads were formally employed, despite not registering a statistically significant increase in overall utilization, also shifted care seeking from private to public facilities following the removal of user fees.

Table 5: Heterogeneous effect of user fee removal on utilization

	Overall Utilization	**Private Utilization**
	(1)	(2)
Education Level of Household Head		
No Education	0.106(0.063)*	0.013(0.021)
Primary	0.062(0.029)**	−0.022(0.009)**
Secondary	0.051(0.037)	−0.034(0.021)
College	0.033(0.049)	−0.031(0.034)
Occupation of Household Head		
Not Employed	0.112(0.050)**	−0.018(0.017)
Self Employed	0.024(0.041)	0.011(0.017)
Farming	0.003(0.042)	−0.077(0.021)***
Formally Employed	0.043(0.038)	−0.045(0.020)**

Note: Table shows the effect of the first wave of the user fee removal policy, T1, on utilization of all facilities (Column 1) and private facilities (Column 2) by educational level and occupational status of the head of household. Each reported coefficient is an estimate of θ_3 in a DD linear probability model (LPM) based on Equation 2. Thus, it is interpreted as a percentage point. Standard errors clustered at the district level are reported in parentheses. All models include a full set of year dummies, regional effects, and region by year interaction effects. Covariates include distance to health facility, household size, sex, age, and age squared of an individual, as well as household head characteristics such as occupational status, marital status, and educational level.
* Significant at 10%; ** Significant at 5%; *** Significant at 1%

5.2 Impact on Medical Spending

The two waves of the removal policy significantly reduced the proportion of individuals incurring any medical spending but increased medical spending for those who continued to incur it, although this increase was not statistically significant (Table 6). As a consequence, both waves of the removal policy did not significantly impact overall medical spending.

The first wave of removals, T1, reduced the likelihood of incurring any spending in the short term by 26.6pp, and this effect reduced slightly to 19.3pp in the long term (Column 1, Table 6). These effects are highly significant at the 1% level. On the other hand, the second wave, T2, reduced the likelihood of incurring any spending by 8.7pp. This effect is significant at the 10% level.

On the other hand, estimates at the intensive margin (GLM part-Column 2) show that the amount of spending (conditional spending) did not change following the re-

moval of user fees in the T1 and T2 groups. In fact, though not statistically significant, there was an upward pressure on the amount of spending. This upward pressure on conditional spending dampened the reduction in spending that resulted from reduced probability of incurring any spending. Thus, unconditional spending was left unchanged (Column 3).

Table 6: Short and long term marginal effects of the removal of user fees on overall medical spending

	Any spending (Probit part)	**Conditional spending (GLM part)**	**Unconditional spending (Probit + GLM)**
	(1)	**(2)**	**(3)**
Short term effect			
T1 × yr2006	−0.266(0.036)***	3.290(14.751)	−6.216(12.653)
Long term effects			
T1 × yr2010	−0.193(0.047)***	17.055(12.546)	2.755(6.598)
T2 × yr2010	−0.087(0.050)*	6.775(10.981)	1.840(5.130)
Pre-treatment effects			
T1 × yr2002	0.005(0.040)	−12.250(24.527)	−3.773(11.093)
T2 × yr2002	0.003(0.052)	−19.567(22.259)	−6.797(10.372)
T2 × yr2006	−0.050(0.046)	0.706(15.928)	−2.249(10.444)
Observations	26,705	13,315	
T1 group baseline mean	0.607	17.177	
T2 group baseline mean	0.580	13.851	
Control group baseline mean	0.706	51.459	

Note: Table shows the effect of the first and second wave of the removal policy, T1 and T2 respectively, on medical spending for all individual who reported being sick or injured. Bootstrap standard errors (1000 replications) clustered at the district level are reported in parentheses. Estimates are from a two-part model where the first part (Column 1) models the probability of incurring any spending (extensive margin) using a probit model of Equation (1) with the 1998 year dummy dropped. The marginal effects of the probit model are computed according to Equation (3). The second part (Column 2) are effects of the policy on the intensive margin–amount of spending conditional on one having incurred any. The amounts are marginal effects computed according to Equation (3) from a GLM regression model (Gamma distribution and log-link) on Equation (1). Column (3) presents the effect of the policy on the whole conditional mean of medical spending (combination of the extensive and intensive margin). The lower panel gives the baseline (2004) mean of each outcome for treated and control groups. All models include a full set of year dummies, regional effects, and region by year interaction effects. Covariates include household size, distance to health facility, sex, age, and age squared of an individual, as well as household head characteristics such as occupational status, marital status, and educational level.
* Significant at 10%; ** Significant at 5%; *** Significant at 1%

All pretreatment interaction effects for extensive (any spending) and intensive (conditional spending) part of medical spending are not statistically different from zero, lending support to the descriptive evidence of common pre-treatment trends in Section 3.2.1.

6 Robustness

6.1 Parallel Trends

Our identification strategy relies on the parallel trends assumption which says that changes in outcomes would have been the same for the treated and control groups in the absence of treatment. To assess the validity of this assumption, we examined trends in utilization and medical spending in Section 3.2.1 and established that our outcomes had similar pre-treatment trends. We then formally tested for common pre-treatment trends by conducting placebo tests were our models included pre-treatment interaction effects. All pre-treatment effects were insignificant, lending support to our identifying assumption. For further testing, we applied a fully flexible difference-in-difference model proposed by Mora and Reggio (2012) and Mora (2015). The common trends assumption was met for all outcomes. Results and details of the fully flexible model is given in Appendix A.2.

6.2 Omitted Variables and Empty models

Although we controlled for individual covariates and included regional fixed effects as well as year by region effects, there may be concerns that other district level time varying characteristics may bias our estimators. To gauge this possibility, we were able to collect data on district level funding. Unfortunately, district funding was not available for 1998, 2002 and 2010. Hence, we could not include this variable in the main analysis. However, the two years, 2004 (before the first intervention) and 2006 (after) could be used to assess the sensitivity of estimates to possible omitted time varying district variables. District funding is perhaps one of the key variables that affects performance, quality of care and availability of essential services and thus utilization as well as individual medical spending. The inclusion of this key district level variable did not change our results (See Table A3 and A4 in the Appendix).

Still on the issue of control variables, the performance of difference-in-difference models can be assessed on how sensitive the estimates are to the addition of covariates. We estimated all models without control variables and with a limited set of control variables and results are presented (Table A5 and A6 in the appendix). Estimates of the effect are broadly similar between models with and without covariates.

6.3 Differential Selection Into Reporting Sickness or Injury

Since utilization is driven by the likelihood of reporting sickness or injury, and vice-versa, there is a possibility of bias where the probability of reporting sickness/injury changes differently between treated and control areas. Table A7 shows that the proportion of individuals reporting sickness in treated and control areas did not change differently. However, the coefficient is significant at the 10% level for the T2 group in 2010. Of course since this particular regression is conducted on the whole dataset (361,336 observations), finding a significant effect at 10% may not necessarily imply bias. It may however also suggest that the removal of user fees led to a slight increase in reporting sickness or injury which is a form of moral hazard.

6.4 Randomization Tests

High correlation in observations over time (serial correlation) or within an areas (intra-cluster correlation) has a potential to underestimate standard errors. Since the user fee removal policy was implemented at the district level one needs to cluster the standard errors at the district level or aggregate the data in order to deal with the problem of within cluster correlation (Bertrand et al., 2004). We clustered standard errors at the district level.

Serial correlation is also a possible threat to the validity of our findings, and we have not addressed it yet. In a difference-in-difference strategy, outcomes are mostly positively correlated over time, and this is worsened by the fact that the treatment indicator itself is highly serially correlated; because it is kept on–equal to one–throughout the study period. This severely underestimates the standard errors of conventional DD models. Using randomly generated placebo interventions, Bertrand et al. (2004) found a false rejection rate, at the 5% level, of up to 44% even after standard errors are clustered at the group level or the data is aggregated to the group level. Since these were placebo interventions, significant effects (rejection rates) were expected in approximately 5% of the simulations.

To assess the extent of this problem, we conducted randomization experiments similar to those of Bertrand et al. (2004). This involves the 54 districts (T1 group) and the urban areas of the 18 districts (control group). If serial correlation is not a problem, the rejection rate in simulations where treatment is turned on in both 2006 and 2010 (call it the pooled model) should be close to the rejection rate where

treatment effect is only turned on in either 2010 or 2006.[7] We found a 6.1% rejection rate in simulation of the pooled model while the other model yielded a rejection rate of 7.1%, suggesting that serial correlation is not a problem in our setup.

6.5 Falsification Tests and Compositional changes

There is still a possibility that other events affected treated and control groups differently and hence generated a false "treatment effect". Although we see no other apparent reason or possible concurrent event that would significantly reduce the percentage of individuals incurring positive spending while at the same time drastically changing utilization, it is important to query this possibility. One way of doing this is to conduct falsification tests were we ask if the removal of user fees significantly affected covariates that are expected to be unresponsive to the removal policy. To achieve this, we fit a simple difference-in-difference models to all our covariates.[8]

In particular, very significant changes in key covariates such as employment status, distance to a health facility, education, especially in the post treatment period, is evidence of other things happening other than removal of user fees. Tables 2 and 3, Column 3-6, in Section 3.2.2 shows that almost all key covariates were not affected differently in the T1^u group compared to the control.

7 Discussion

Providing free health care for the poor such as removing user fees may increase utilization of health services and reduce the occurrence as well as severity of health shocks. This is especially true if the increases in utilization arise out of genuine need and are sustained over time. Removal of user fees, by reducing medical spending, may also

[7] In this experiment, for the pooled model, we randomly select 54 districts (regardless of treatment status) and designate them as "treatment" and let them keep the treatment status in 2010. We then record the treatment effect and standard error. We conduct this procedure 1,000 times each time taking note of the standard errors. In an alternative model, we randomly draw 54 districts and designate them as treated. In this case, however, treatment is turned on in either 2006 or 2010 but not both years.

[8] For example, to check for differential changes between T1 and control as reported in Table 2, we a difference-in-difference model of the form:

$$\begin{aligned} X_{cidt} &= \phi_0 + \phi_1 \text{yr1998} + \phi_2 \text{yr2002} + \phi_3 \text{yr2006} + \phi_4 \text{yr2010} + \phi_5 \text{T1}_d \\ &+ \phi_6(\text{T1}_d \times \text{yr1998}) + \phi_7(\text{T1}_d \times \text{yr2002}) + \phi_8(\text{T1}_d \times \text{yr2006}) + \phi_9(\text{T1}_d \times \text{yr2010}) + \epsilon_{ijt} \end{aligned}$$

where X_{cidt} is covariate c for individual i residing in district j at time t and $\text{T1}_d = 1$ if district j was treated. We report p-values of the interaction effects ϕ_6,ϕ_7, ϕ_8 and ϕ_9 in Table 2 and Table 3 for T2.

free up resources which would then be available for other competing household welfare enhancing needs.

In this paper, we examine the extent to which provision of free care in public health facilities changes utilization and medical spending. Using a natural experiment accorded by the stepwise removal of user fees and five waves of rich nationally representative household survey day, we address a number of questions that surround the user fees removal debate. Does the removal of user fees in public health facilities increase overall utilization of health services? In addition to the increase in overall utilization, are individuals induced to switch from private to public facilities as a result of user fee removal in public facilities? Are there heterogeneities in how utilization of individuals of different socioeconomic background responds? To what extent does medical spending respond to the removal of user fees in public health facilities. Are the changes observed in the short term sustained over the long term?

We find that removal of user fees increased overall use of health services in Zambia.Interestingly, the effects of the user fee removal policy remained strong in the long term effect, 48 months (four years) after the removal. This means that concerns that changes in utilization following removal of user fees are temporal, and not sustained in the long term (Lagarde and Palmer, 2008), are not supported by this study.

While previous studies have also shown that removal of user fees in Zambia increased utilization of public health facilities (Masiye et al., 2010; Lagarde et al., 2012; Onde, 2009), we demonstrate that this increase was generated by two mechanisms, namely, overall increase in the use of health services (uptake effect) and shifting care seeking from private to public facilities (switching effect). Some randomized experiments of free health care have shown that removal of user fees or providing health insurance may only generate a switching effect without a change in the overall use of health services (Powell-Jackson et al., 2015; Levine et al., 2016). Although we find evidence of switching in our natural experiment, our findings show that the increase in overall use of health service (uptake effect) was larger, and much stronger, than the switching effect. To the best of our knowledge, our study is the first to separate the magnitude of uptake and switching effects for a complex policy implemented at national scale. Mwabu et al. (1995) conducted a descriptive study of the effect of the temporal suspension of user fees in two districts of Kenya and found that utilization of public facilities increased while that of private facilities declined. This analysis was however unable to tease out the magnitude of the increase in overall use of health services. Moreover, since it was a before and after study without a control group, it

is hard to determine whether or not the observed changes in utilization patterns were due to the removal of user fees or other concurrent factors.

Our study also documents heterogeneities in the utilization response. The user fee removal policy led to an increase in the uptake of health services for individuals of low socioeconomic background. This is encouraging given that the overriding objective of the user fee removal policy is to increase access for individual from low socioeconomic background. However, we find evidence that individuals from relatively higher socioeconomic backgrounds shifted care seeking from private to public facilities.

As expected, we find evidence that user fees reduced the proportion of people incurring any medical spending. However, those who were still incurring any spending experienced an economically relevant, albeit statistically insignificant, increase in medical spending. The net effect is that the removal policy did not impact unconditional medical spending significantly. This result is collaborated by findings from Uganda (Xu et al., 2006; Nabyonga Orem et al., 2011). Evidence from within Zambia also show that medical spending in primary health care facilities may not have responded to the removal of user fees as expected (Masiye et al., 2016; Hadley, 2011). This may be attributed to the fact that public health facilities may not have been able to provide complete services such as drugs, medical examinations, etc, to match the increased demand so that individuals had to rely on the private markets for these services (Masiye et al., 2010; Hadley, 2011; Xu et al., 2006; Nabyonga Orem et al., 2011; Ministry of Health, 2007; Masiye et al., 2010; Onde, 2009). There is also evidence that some additional charges replaced user fees, formal and informal. For example, Hadley (2011) found that there were charges to have children weighed at the facility, informal medical insurance arrangements, and book fees for maintaining medical records, among others. The study also found that even when drugs were available in health facilities, patients would be given prescriptions to buy drugs from private drug stores, most of which were owned by health facility workers. It is therefore not surprising that even after the nation-wide removal of user fees in 2012, a national health expenditure survey documented significant medical spending for individual visiting primary health care facilities (Masiye et al., 2016).[9]. The finding that some form of free care policy, while increasing utilization, may not be very effective in reducing total medical spending is

[9]That total medical spending was unchanged is also consistent with the National Health Accounts (NHA) for Zambia which reviewed that the share of Household Health Expenditure (HHE) in Total Health Expenditure (THE) was 27% in 2006, from 26.8 in 2005 (Ministry of Health, 2009). In fact, NHA data shows that out of pockets expenditure (OPE) as share of total health spending significantly increased immediately after the removal of user fees and remained high thereafter (World Health Organization, 2015)

also consistent with the health insurance literature (Liu and Zhao, 2014; Wagstaff and Lindelow, 2008; Fink et al., 2013; Wagstaff et al., 2009; Nguyen, 2012; Ataguba and Goudge, 2012).

Limitations

The key limitation of our study is the non-availability of GPS information to exactly classify which households could have been treated in the second wave of the user fee removal policy, T2. Despite the fact that our classification performs extremely well, the long term estimate of T1 and the effect of T2 is lower bound if there was any misclassification.

Secondly, our measure of medical spending does not fully characterize health expenditure because Living Conditions Monitoring Surveys only collects spending on consultation, purchase of drugs and examinations. This leaves out other significant health care related costs such as transportation which in some cases are more significant costs than payments at the point of care.

8 Conclusion

Consensus appears to have emerged that removing user fees is an effective strategy for improving Universal Health Coverage (UHC). The consensus draws heavily on studies that report dramatic increases in utilization. A number of low and middle income countries have since removed user fees while others are considering to do so. However, there are are a number of unresolved question regarding the exact impact of user fee removal policy. Addressing these questions is important in providing policy makers with information that they require in their search for more efficient health care financing systems which ensure equitable access to care and provide financial protection to the population.

Most studies so far have only demonstrated increases in utilization of public facilities. We do not know whether or not this is a result of changes in overall rates of care seeking or switching from private to public facilities. Similarly, although the removal of user fees is heralded as important in increasing access for the poor, the distributional or heterogeneous effects of the removal policy is unclear. It is not clear whether the policy is successful in increasing utilization more for individuals from low socioeconomic backgrounds. In addition, most studies only assess changes in utilization without looking at whether medical spending reduces, which in itself is an end of

the health system. There is also no clear evidence on whether changes observed in the short term are sustained in the long term.

To fill the knowledge gap in the literature, we exploit the step-wise implementation of removal of user fees in Zambia to create treatment and control groups, and use large nationally representative surveys to estimates effects in a difference-in-difference framework.

We find that the two waves of the user fee removal policy significantly increased overall rates of care seeking and moderately led to switching from private to public facilities in both the short and long term. Overall rates of care seeking increased more for individuals from lower socioeconomic backgrounds. Despite the fact that the removal policy reduced the financial barrier to access by substantially reducing the proportion of individual incurring any spending (extensive margin), and thus significantly increasing utilization, it did not have a significant effect on conditional spending (intensive margin). The intensive and extensive margin worked in opposite directions to leave total (unconditional) medical spending unchanged.

References

Acemoglu, D. (2010). Theory, general equilibrium, and political economy in development economics. *Journal of Economic Perspectives*, 24(3):17–32.

Ai, C. and Norton, E. C. (2003). Interaction terms in logit and probit models. *Economics Letters*, 80(1):123–129.

Asfaw, A., von Braun, J., and Klasen, S. (2004). How big is the crowding-out effect of user fees in the rural areas of Ethiopia? implications for equity and resources mobilization. *World Development*, 32(12):2065–2081.

Ashraf, N., Fink, G., and Weil, D. N. (2014). Evaluating the effects of large scale health interventions in developing countries: The Zambian Malaria Initiative. In Edwards, S., Johnson, S., and Weil, D. N., editors, *African Successes: Human Capital, Volume 2*. University of Chicago Press.

Ataguba, J. E.-O. and Goudge, J. (2012). The impact of health insurance on healthcare utilisation and out-of-pocket payments in South Africa. *Geneva Pap R I-Iss P*, 37(4):633–654.

Barber, S., Bonnet, F., and Bekedam, H. (2004). Formalizing under-the-table payments to control out-of-pocket hospital expenditures in Cambodia. *Health Policy and Planning*, 19(4):199–208.

Basu, S., Andrews, J., Kishore, S., Panjabi, R., and Stuckler, D. (2012). Comparative performance of private and public healthcare systems in low-and middle-income countries: a systematic review. *PLoS med*, 9(6):e1001244.

Beegle, K., Weerdt, J. D., Friedman, J., and Gibson, J. (2012). Methods of household consumption measurement through surveys: Experimental results from Tanzania. *Journal of Development Economics*, 98(1):3–18. Symposium on Measurement and Survey Design.

Bertrand, M., Duflo, E., and Mullainathan, S. (2004). How much should we trust differences-in-differences estimates? *The Quarterly Journal of Economics*, 119(1):249–275.

Bitran, R. (2014). Universal health coverage and the challenge of informal employment : lessons from developing countries. Health, nutrition, and population (hnp) discussion paper, World Bank Group, Washington DC.

Blas, E. and Limbambala, M. (2001). User-payment, decentralization and health service utilization in Zambia. *Health Policy and Planning*, 16(suppl 2):19–28.

Cai, L., Mavromaras, K., and Oguzoglu, U. (2014). The effects of health status and health shocks on hours worked. *Health Economics*, 23(5):516–528.

Cohen, J. and Dupas, P. (2010). Free distribution or cost-sharing? evidence from a randomized malaria prevention experiment. *The Quarterly Journal of Economics*, 125(1):1–45.

CSO (2011). Living Conditions Monitoring Survey Report 2006. Survey report, Central Statistical Office, Government of the Republic of Zambia, Lusaka, Zambia.

CSO (2012). Living Conditions Monitoring Survey Report 2006 and 2010. Survey report, Central Statistical Office, Government of the Republic of Zambia, Lusaka, Zambia.

Dhanaraj, S. (2016). Effects of parental health shocks on children's schooling: Evidence from Andhra Pradesh, India. *International Journal of Educational Development*, 49:115–125.

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

Dzakpasu, S., Powell-Jackson, T., and Campbell, O. M. (2014). Impact of user fees on maternal health service utilization and related health outcomes: a systematic review. *Health Policy and Planning*, 29(2):137–150.

Falkingham, J. (2004). Poverty, out-of-pocket payments and access to health care: evidence from Tajikistan. *Social Science & Medicine*, 58(2):247 –258. Adjusting for Market Failure: Challenges in Public Health Alternatives.

Fink, G., Robyn, P. J., Sié, A., and Sauerborn, R. (2013). Does health insurance improve health?: Evidence from a randomized community-based insurance rollout in rural Burkina Faso. *Journal of Health Economics*, 32(6):1043–1056.

Frondel, M. and Vance, C. (2013). On Interaction Effects: The Case of Heckit and Two-Part Models. *Journal of Economics and Statistics*, 233(1):22 –38.

Gertler, P. and Gruber, J. (2002). Insuring consumption against illness. *American Economic Review*.

Godlonton, S. and Okeke, E. N. (2016). Does a ban on informal health providers save lives? evidence from Malawi. *Journal of Development Economics*, 118:112–132.

Hadley, M. (2011). Does increase in utilisation rates alone indicate the success of a user fee removal policy? a qualitative case study from Zambia. *Health Policy*, 103(2–3):244–254.

Institute for Health Metrics and Evaluation, Human Development Network, The World Bank (2013). The Global Burden of Disease: Generating Evidence, Guiding Policy — Sub-Saharan Africa Regional Edition. Policy Report, IHME, Seattle, WA.

Islam, A. and Maitra, P. (2012). Health shocks and consumption smoothing in rural households: Does microcredit have a role to play? *Journal of Development Economics*, 97(2):232 – 243.

James, C., Hanson, K., McPake, B., Balabanova, D., Gwatkin, D., Hopwood, I., Kirunga, C., Knippenberg, R., Meessen, B., Morris, S., Preker, A., Souteyr, Y., Tibouti, A., Villeneuve, P., and Xu, K. (2006). To retain or remove user fees? *Applied Health Economics and Health Policy*, 5(3):137–153.

Jolliffe, D. (2001). Measuring absolute and relative poverty: the sensitivity of estimated household consumption to survey design. *Journal of Economic and Social Measurement*, 27(1).

Kremer, M. and Miguel, E. (2007). The illusion of sustainability. *The Quarterly Journal of Economics*, 122(3):1007–1065.

Lagarde, M., Barroy, H., and Palmer, N. (2012). Assessing the effects of removing user fees in Zambia and Niger. *Journal of Health Services Research & Policy*, 17(1):30–36.

Lagarde, M. and Palmer, N. (2008). The impact of user fees on health service utilization in low- and middle-income countries: how strong is the evidence? *Bulletin of the World Health Organization*, 86:839 – 848C.

Levine, D., Polimeni, R., and Ramage, I. (2016). Insuring health or insuring wealth? an experimental evaluation of health insurance in rural Cambodia. *Journal of Development Economics*, 119:1–15.

Lim, S. S., Stein, D. B., Charrow, A., and Murray, C. J. (2008). Tracking progress towards universal childhood immunisation and the impact of global initiatives: a

systematic analysis of three-dose diphtheria, tetanus, and pertussis immunisation coverage. *The Lancet*, 372(9655):2031–2046.

Lindkvist, I. (2013). Informal payments and health worker effort: a quantitative study from Tanzania. *Health Economics*, 22(10):1250–1271.

Litvack, J. I. and Bodart, C. (1993). User fees plus quality equals improved access to health care: Results of a field experiment in Cameroon. *Social Science & Medicine*, 37(3):369 –383.

Liu, H. and Zhao, Z. (2014). Does health insurance matter? Evidence from China's urban resident basic medical insurance. *Journal of Comparative Economics*, 42(4):1007–1020.

Masiye, F., Chitah, B. M., and McIntyre, D. (2010). From targeted exemptions to user fee abolition in health care: Experience from rural Zambia. *Social Science & Medicine*, 71(4):743 –750.

Masiye, F., Kaonga, O., and Kirigia, J. M. (2016). Does user fee removal policy provide financial protection from catastrophic health care payments? evidence from Zambia. *PLoS ONE*, 11(1):1–15.

Meessen, B., Damme, W. V., Tashobya, C. K., and Tibouti, A. (2007). Poverty and user fees for public health care in low-income countries: lessons from Uganda and Cambodia. *The Lancet*, 368(9554):2253 –2257.

Meessen, B., Hercot, D., Noirhomme, M., Ridde, V., Tibouti, A., Tashobya, C. K., and Gilson, L. (2011). Removing user fees in the health sector: a review of policy processes in six sub-Saharan African countries. *Health Policy and Planning*, 26(suppl 2):ii16–ii29.

Ministry of Health (2007). Health Sector Joint Annual Review Report, 2006 . Technical report, Ministry of Health, Government of the Republic of Zambia, Lusaka, Zambia.

Ministry of Health (2009). Zambia National Health Accounts: 2003-2006. Final report, Ministry of Health, GRZ, Lusaka, Zambia.

MoH (2007). Revised Guidelines on the Removal of User Fees in Government and Mission Health Facilities in Zambia . Technical report, Ministry of Health, Government of the Republic of Zambia, Lusaka, Zambia.

Mora, R. (2015). didq: A command for treatment-effect estimation under alternative assumptions. *Stata Journal*, 15(3):796–808(13).

Mora, R. and Reggio, I. (2012). Treatment effect identification using alternative parallel assumptions. Working Paper 12-33, Economic series (48), Departamento de Economía, Universidad Carlos III de Madrid.

Mwabu, G., Mwanzia, J., and Liambila, W. (1995). User charges in government health facilities in Kenya: effect on attendance and revenue. *Health Policy and Planning*, 10(2):164–170.

Nabyonga Orem, J., Mapunda, M., Musango, L., and Mugisha, F. (2013). Long-term effects of the abolition of user fees in Uganda. African health monitor paper, World Health Organization.

Nabyonga Orem, J., Mugisha, F., Kirunga, C., Macq, J., and Criel, B. (2011). Abolition of user fees: the Uganda paradox. *Health Policy and Planning*, 26(suppl 2):ii41–ii51.

Nguyen, C. V. (2012). The impact of voluntary health insurance on health care utilization and out-of-pocket payments: new evidence for Vietnam. *Health Economics*, 21(8):946–966.

Onde, R. (2009). Impact of abolishment of user fees in rural health centres: A case of Chongwe district. Master's thesis, School of Business, Copperbelt University, Kitwe, Zambia.

Powell-Jackson, T., Hanson, K., Whitty, C. J., and Ansah, E. K. (2014). Who benefits from free healthcare? evidence from a randomized experiment in Ghana. *Journal of Development Economics*, 107:305–319.

Powell-Jackson, T., Macleod, D., Benova, L., Lynch, C., and Campbell, O. M. R. (2015). The role of the private sector in the provision of antenatal care: a study of demographic and health surveys from 46 low- and middle-income countries. *Tropical Medicine & International Health*, 20(2):230–239.

Pradhan, M. (2009). Welfare analysis with a proxy consumption measure: Evidence from a repeated experiment in Indonesia. *Fiscal Studies*, 30(3–4):391–417.

Ridde, V. and Haddad, S. (2009). Abolishing user fees in Africa. *PLoS Med*, 6(1):e1000008.

Sandefur, J. and Glassman, A. (2015). The Political Economy of Bad Data: Evidence from African Survey and Administrative Statistics. *The Journal of Development Studies*, 51(2):116–132.

Soucat, A., Gandaho, T., Levy-Bruhl, D., de Bethune, X., Alihonou, E., Ortiz, C., Gbedonou, P., Adovohekpe, P., Camara, O., Ndiaye, J.-M., Dieng, B., and Knippenberg, R. (1997). Health seeking behaviour and household health expenditures in Benin and Guinea: the equity implications of the Bamako Initiative. *The International Journal of Health Planning and Management*, 12(S1):S137–S163.

Tanaka, S. (2014). Does abolishing user fees lead to improved health status? evidence from post-apartheid South Africa. *American Economic Journal: Economic Policy*, 6(3):282–312.

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

UNDP (2014). Human Development Report 2014 :Reducing Vulnerabilities and Building Resilience. Health, nutrition, and population (hnp) discussion paper, United Nations Development Program, New York, USA.

Wagstaff, A. (2010). Social health insurance reexamined. *Health Economics*, 19(5):503–517.

Wagstaff, A. and Lindelow, M. (2008). Can insurance increase financial risk?: The curious case of health insurance in China. *Journal of Health Economics*, 27(4):990–1005.

Wagstaff, A., Lindelow, M., Jun, G., Ling, X., and Juncheng, Q. (2009). Extending health insurance to the rural population: An impact evaluation of China's new cooperative medical scheme. *Journal of Health Economics*, 28(1):1–19.

World Bank (2015). World Development Indicators 2015. doi:10.1596/978–1-4648–0440–3. License: Creative Commons Attribution CC BY 3.0 IGO, World Bank, Washington DC.

World Health Organization (2010). The World Health Report-Health Systems Financing: The Path to Universal Coverage. Technical report, World Health Organization, Geneva, Switzerland.

World Health Organization (2015). WHO Global Health Expenditure Database. Technical report. http://apps.who.int/nha/database/Select/Indicators/en. Accessed:2015-06-28, type=National Health Accounts,.

Xu, K., Evans, D. B., Kadama, P., Nabyonga, J., Ogwal, P. O., Nabukhonzo, P., and Aguilar, A. M. (2006). Understanding the impact of eliminating user fees: Utilization and catastrophic health expenditures in Uganda. *Social Science & Medicine*, 62(4):866 – 876.

A Appendix

A.1 Assessing Treatment Classification for T2

The second wave of the removal policy, T2, extended the removal of user fees to the T2 group–rural areas of the remaining 18 districts. However, the rural/urban classification of areas within these 18 districts in our data is not exactly the same as defined in the removal policy. The classification in the data uses a more detailed criteria while the removal policy defined rural areas only on the basis of radius distance from the district administration centers of each of the 18 districts. Specifically, user fees were removed in all primary health facilities located outside a radius of 15 or 20 km from the districts centers, depending on localization relative to the line of rail. Ideally, we would need GPS coordinates for both households and facilities to precisely determine which household is actually located in a treated area. However, GPS coordinates were not collected in the surveys.

Since distances in the removal policy were determined in such a way that health facilities exempted from user fees would serve rural households, we use the rural classification from the Central Statistical Office to define households as treated (belonging to the T2 group) if it is located in an enumeration area that is classified as rural. The classification of enumeration areas as either rural or urban is based on population size, economic activity (agricultural or not), and presence of basic modern facilities such as piped water, tarred roads, post office and other services (CSO, 2012).

To assess how well this rural/urban definitions in the data performs in distinguishing treated from non-treated individuals, we check how the proportion of individuals incurring any medical spending changed between rural areas (supposedly treated-T2 group) and urban areas (supposedly controls). It is expected that after the removal of user fees in 2007, this measure falls significantly in areas that are supposedly treated, while it should be non-decreasing in areas deemed as controls. The data shows that while the proportion of individuals incurring any spending in the supposedly treated areas fell by almost 20 percentage points (pp) between 2006 and 2010, it was almost unchanged in the supposed control (Table A1, Panel A4).

Table A1: Trends in medical spending in public health facilities

	T2 Group				Control Group			
Variable	**2002**	**2004**	**2006**	**2010**	**2002**	**2004**	**2006**	**2010**
Incurred any spending (%)	60.75	82.57	59.38	40.17	69.08	80.99	79.45	81.17
Conditional spending	8.35	12.55	12.85	7.17	42.68	43.10	45.53	30.68
Number of observations	873	701	673	1,032	1,969	1,972	1,586	2,391

Note: Table shows means medical spending for the T2 group and control. Note that, as opposed to computing spending for all individuals who reported being sick or injured as we did in the main analysis, medical spending here is conditional on utilizing public health facilities. Conditional Health Spending is in Zambian Kwacha (ZMW) at 2010 prices. The ZMW/USD exchange rate in 2010 was 5 Kwacha per USD.
* Significant at 10%; ** Significant at 5%; *** Significant at 1%

A.2 Fully Flexible Model

The model proposed by Mora and Reggio (2012) allows provides an alternative way of estimating the treatment effect of the removal policy while at the same time testing for the parallel trends assumption. As opposed to estimating one treatment effect of an intervention for each post-treatment data point (e.g 2006 or 2010)–irrespective of the number of pre-treatment data points–the model proposed by Mora and Reggio (2012) allows one to estimate r treatment effects for each post-treatment data point, where r is the number of pre-treatment data points and each effect, β_r, is estimated under a different parallel assumption. Only when treatment effects under different parallel assumptions are equivalent is the parallel trends assumption met, and our assumption of common trends valid.

The advantage with the method proposed by Mora and Reggio (2012) is that even if the estimates from one post-treatment data point, e.g for T1 $\times$ 2006, under different parallel assumptions are not equivalent, one is able to check how the effect varies, in both magnitude and significance across different parallel assumptions. Actually, Mora and Reggio (2012) show that the practice of allowing for flexibility in difference-in-difference models by including terms such as linear or quadratic trends imply different identifying assumptions from the ones the authors refer to, so that the reported effects may not be the true treatment effects.[10]

We focus on the first wave of removals, T1, and estimate the fully flexible model on three outcomes, namely, overall (any) utilization, whether an individual incurred any spending or not, and the amount of spending conditional on spending. Given that

[10]They applied their method to papers published in top economics journals and find that when flexible dynamics are applied, the significance of the results is affected in 6 of the 13 cases and only 3 in 9 cases was the common trends assumption satisfied.

we have three (two) pre-treatment periods for overall utilization (medical spending), we are able to estimate three (two) short term and three (two) long term effects of T1. For example. the computation of the three short term and three long term effect for overall utilization follows three stages. Firstly, the first set of short and long term effect are computed using the baseline data (2004) only. The identifying assumption in this case is the parallel paths assumption, which Mora and Reggio (2012) term parallel-1. Then the second set of short term and long term effects are estimated using both the 2004 and 2002 pre-treatment data. The assumption here is parallel growths, called parallel-2 and it is equivalent to allowing for linear trends in a difference in difference model. The last set of effects are computed using all the pre-treatment datasets (1998,2002, and 2004). This is called parallel-3 and its equivalent to allowing for quadratic trends in a difference in difference model.

We make use of the Stata package by Mora (2015) and focus on T1 which gives the short and long term effects. The models are estimated using ordinary least squares (OLS). Thus utilization and the first part of the two part model of expenditure applies a linear probability model (LPM) while the second part is OLS on log transformed expenditure.

Estimation results are presented in Table A2. The results are broadly consistent with our finding in the main analysis although the estimates and standard errors are slightly larger. Panel B shows test statistics and p-values for of the parallel trends assumption test and the equivalence of short and long term effects. The common trends assumption was met for all outcomes, i.e. we do not reject H_0 of common trends at the 10% level. The tests also reveals that there was not statistically significant difference between short and long term effects

Table A2: Short and long term effects of the removal of user fees: Fully flexible model (Mora, 2015)

PANEL A: Short and long term effect under different parallel assumptions

	Overall utilization			Prob. of medical spending		Conditional spending	
	Parallel-1	**Parallel-2**	**Parallel-3**	**Parallel-1**	**Parallel-2**	**Parallel-1**	**Parallel-2**
T1 × yr2006	0.098***	0.108*	0.090	−0.236***	−0.258***	0.060	−0.133
	(0.034)	(0.057)	(0.109)	(0.035)	(0.060)	(0.120)	(0.569)
T1 × yr2010	0.102**	0.119	0.066	−0.240***	−0.286***	0.019	−0.369
	(0.043)	(0.096)	(0.270)	(0.036)	(0.102)	(0.145)	(0.388)
Observations	28,232	28,232	28,232	23,783	23,783	11,892	11,892

PANEL B: Test for Common Pre-treatment trends and equality of short and long term effects

	Overall utilization		Prob. of medical spending		Conditional spending	
	F	**p-value**	**F**	**p-value**	**F**	**p-value**
Test for pre-treatment trends						
H_0: Common Trends	1.051	.5913	0.2805	0.5964	1.772	0.183
Test of equality of short and long term effects						
Under Parallel-1						
H_0: T1×yr2006=T1×yr2010	0.018	0.892	0.024	0.877	0.103	0.748
Under Parallel-2						
H_0: T1×yr2006=T1×yr2010	0.067	0.795	0.230	0.631	1.356	0.244
Under Parallel-3						
H_0: T1×yr2006=T1×yr2010	0.021	0.885				

Note: The table shows OLS estimates in a fully flexible DD model for the T1 group and the control. The short and long term effect under parallel-1 assumes parallel paths and uses only 1 pre-treatment period (2004). The effects under parallel-2 uses the 2004 and 2002 pre-treatment data set and assumes parallel growths in outcomes. Under parallel-3, all pre-treatment data is used (2004,2002 & 1998). Parrellel-3 assumes quadratic growth in outcomes. The first test is a test of common trends which is equivalent to a test that the 3 (or 2 under medical spending) parallel path assumptions are equal. The last three tests checks whether the short term and long term effects are equal under each parallel assumption. Robust Standard errors clustered at the district level are reported in brackets.
* Significant at 10%; ** Significant at 5%; *** Significant at 1%

Controlling for district funding

Table A3: Short term effect of user fee removal on overall, public and private facility utilization

	Overall Utilization	**Public Utilization**	**Private Utilization**
	(1)	(2)	(3)
T1 × yr2006	$0.079(0.033)^{**}$	$0.120(0.032)^{***}$	$-0.041(0.019)^{**}$
Observations	11,402	11,402	11,402
T1 group baseline mean	0.521	0.497	0.024
Control group baseline mean	0.575	0.472	0.103

Note: Table shows the effect of the first wave of the removal policy, T1, on utilization of all facilities (Column 1), public facilities (Column 2) and private facilities (Column 3) controlling for district level funding. Funding data was only available for 2004 and 2006. Thus, estimates are only based on the 2004 and 2006 LCMS surveys. Reported coefficients are estimates from DD linear probability models (LPM) corresponding to Equation (1) and are interpreted as percentage points. Standard errors clustered at the district level are reported in parentheses. The lower panel give the baseline (2004) mean of each outcome for treated and control groups. All models include a full set of year dummies, regional effects, and region by year interaction effects. Covariates include district level funding, distance to health facility, household size, sex, age, and age squared of an individual, as well as household head characteristics such as occupational status, marital status, and educational level.
* Significant at 10%; ** Significant at 5%; *** Significant at 1%

Table A4: Short and long term marginal effects of the removal of user fees on overall medical spending

	Any spending (Probit part)	**Conditional spending (GLM part)**	**Unconditional spending (Probit + GLM)**
	(1)	**(2)**	**(3)**
T1 × yr2006	$-0.265(0.037)^{***}$	2.946(11.662)	−5.355(9.991)
Observations	11,402	6,360	
T1 group baseline mean	0.607	17.177	
Control group baseline mean	0.706	51.459	

Table shows the effect of the first wave of the removal policy, T1, on medical spending for all individual who reported being sick or injured controlling for district level funding. Funding data was only available for 2004 and 2006. Thus, estimates are only based on the 2004 and 2006 LCMS surveys. Bootstrap standard errors (1000 replications) clustered at the district level in parentheses. Estimates are from a two-part model where the first part (Column 1) models the probability of incurring any spending (extensive margin) using a probit model of Equation (1) with the 1998 year dummy dropped. The marginal effects of the probit model are computed according to Equation (3). The second part (Column 2) are effects of the policy on the intensive margin–amount of spending conditional on one having incurred any. The amounts are marginal effects computed according to Equation (3) from a GLM regression model (Gamma distribution and log-link) on Equation (1). Column (3) presents the effect of the policy on the whole conditional mean of medical spending (combination of the extensive and intensive margin). The lower panel gives the baseline (2004) mean of each outcome for treated and control groups. All models include a full set of year dummies, regional effects, and region by year interaction effects. Covariates include household size, distance to health facility, sex, age, and age squared of an individual, as well as household head characteristics such as occupational status, marital status, and educational level.
* Significant at 10%; ** Significant at 5%; *** Significant at 1%

Models without Covariates

Table A5: Short and long term effect of user fee removal on provider choice

	Overall Utilization (1)	**Public Utilization** (2)	**Private Utilization** (3)
Short term effect			
T1 × yr2006	0.077(0.033)**	0.092(0.037)**	−0.015(0.021)
Long term effects			
T1 × yr2010	0.10.8(0.042)**	0.113(0.040)***	−0.004(0.016)
T2 × yr2010	0.154(0.047)***	0.159(0.025)***	−0.005(0.015)
Pre-treatment			
T1 × yr1998	0.021(0.039)	0.012(0.031)	0.009(0.023)
T2 × yr1998	0.027(0.059)	0.012(0.056)	0.015(0.027)
T1 × yr2002	0.032(0.032)	0.029(0.035)	0.004(0.015)
T2 × yr2002	0.033(0.052)	0.034(0.050)	−0.001(0.014)
T2 × yr2006	0.072(0.062)	0.072(0.060)	−0.001(0.020)
Observations	36,319	36,319	36,319
T1 group baseline mean	0.521	0.497	0.024
T2 group baseline mean	0.447	0.433	0.014
Control group baseline mean	0.575	0.472	0.103

Note: Estimates from a difference in difference linear probability model (LPM) corresponding to Equation (1). Estimates are interpreted as percentage points. Standard errors clustered at the district level are reported in parentheses. All models are conditional on reporting sickness or injury in the two weeks prior to the survey. All models include a full set of year dummies, regional effects, and region by year interaction effects. Covariates include distance to health facility, household size, sex, age, and age squared of an individual, as well as household head characteristics such as occupational status, marital status, and educational level.
* Significant at 10%; ** Significant at 5%; *** Significant at 1%

Table A6: Short and long term marginal effects of the removal of user fees on medical spendings

	Any spending (Probit part)	**Conditional spending (GLM part)**	**Unconditional spending (Probit + GLM)**
	(1)	**(2)**	**(3)**
Short term effect			
T1 × yr2006	−0.223(0.032)***	5.080(12.340)	−1.824(6.487)
Long term effects			
T1 × yr2010	−0.186(0.036)***	37.494(17.959)**	12.011(6.472)*
T2 × yr2010	−0.070(0.050)	10.905(17.075)	9.144(7.563)
Pre-treatment effects			
T1 × yr2002	−0.025(0.041)	4.747(13.625)	8.663(7.118)
T2 × yr2002	−0.001(0.047)	0.262(14.622)	8.676(7.822)
T2 × yr2006	−0.035(0.038)	6.705(14.844)	2.423(6.904)
Observations	31,039	15,175	
T1 group baseline mean	0.607	17.177	
T2 group baseline mean	0.580	13.851	
Control group baseline mean	0.706	51.459	

Note: Table shows marginal effects from a DD two part model and the combined effect of the two parts. All amounts are in Zambian Kwacha at 2010 prices. Bootstrap standard errors (1000 replications) clustered at the district level are reported in parentheses. (1) presents probit model marginal effects at the extensive margin of medical spending (whether one incurred any spending or not). In (2), marginal effect at the intensive margin (those with positive spending) are presented. (3) presents the effect of the policy on the whole conditional mean of medical spending (combination of the extensive and intensive margin). All models include a full set of year dummies, regional effects, and region by year interaction effects. Covariates include household size, distance to health facility, sex, age, and age squared of an individual, as well as household head characteristics such as occupational status, marital status, and educational level.
* Significant at 10%; ** Significant at 5%; *** Significant at 1%

Changes in the likelihood of reporting sickness or injury

Table A7: Differences in the likelihood of reporting sickness between the treated (T1 & T2) and the control group

T1×yr1998	T1×yr2002	T1×yr2006	T1×yr2010	T2×yr1998	T2×yr2002	T2×yr2006	T2×yr2010
−0.015	−0.004	0.000	0.010	0.004	0.020	0.021	0.039*
(0.013)	(0.012)	(0.008)	(0.010)	(0.012)	(0.013)	(0.013)	(0.023)
Observations							361,336

Note: Estimates from a difference-in-difference linear probability model (LPM) corresponding to Equation (1). Estimates are interpreted as percentage points. Standard errors clustered at the district level are reported in parentheses. The model includes a full set of year dummies, regional effects, and region by year interaction effects. Covariates include , household size, distance to health facility, sex, age, and age squared of an individual, as well as household head characteristics such as occupational status, marital status, and educational level.
* Significant at 10%; ** Significant at 5%; *** Significant at 1%

PAPER III

Explaining changes in child health inequality in the run up to the 2015 Millennium Development Goals (MDGs): The case of Zambia

Authors: Peter Hangoma, Arild Aakvik, and Bjarne Robberstad.

Under Resubmission: PLOSE ONE

III

Explaining changes in child health inequality in the run up to the 2015 Millennium Development Goals (MDGs): The case of Zambia

Peter Hangoma[a,b*], Arild Aakvik[c], and Bjarne Robberstad[a,d]

a. Centre for International Health, University of Bergen
b. Unit of Health Economics, School of Public Health, University of Zambia
c. Department of Economics, University of Bergen
d. Centre for Intervention Science in Maternal and Child Health, University of Bergen

*Corresponding author

Email: peterhangoma555@gmail.com

Abstract

Background

Child health interventions were drastically scaled up in the period leading up to 2015 as countries aimed at meeting the 2015 target of the Millennium Development Goals (MDGs). MDGs were defined in terms of achieving improvements in average health. Significant improvements in average child health are documented, but evidence also points to rising inequality. It is important to investigate factors that drive the increasing disparities in order to inform the post-2015 development agenda of reducing inequality, as captured in the Sustainable Development Goals (SDGs). We investigated changes in socioeconomic inequality in stunting and fever in Zambia in 2007 and 2014. Unlike the huge literature that seeks to quantify the contribution of different determinants on the observed inequality at any given time, we quantify determinants of changes in inequality.

Methods

Data from the 2007 and 2014 waves of the Zambia Demographic and Health Survey (DHS) were utilized. Our sample consisted of children aged 0–5 years (n = 5,616 in 2007 and n = 12,714 in 2014). We employed multilevel models to assess the determinants of stunting and fever, which are two important child health indicators. The concentration index (CI) was used to measure the magnitude of inequality. Changes in inequality of stunting and fever were investigated using Oaxaca-type decomposition of the CI. In this approach, the change in the CI for stunting/fever is decomposed into changes in CI for each determinant and changes in the effect—measured as an elasticity—of each determinant on stunting/fever.

Results

While average rates of stunting reduced in 2014 socioeconomic inequality in stunting increased significantly. Inequality in fever incidence also increased significantly, but average rates of fever did not reduce.

The increase in the inequality (CI) of determinants accounted for the largest part (42.5%) of the increase in inequality of stunting, while the increase in the effect of determinants explained 35% of the increase. The determinants with the greatest total contribution (change in CI plus change in effect) to the increase in inequality of stunting were mother's height and weight, wealth, birth order, facility delivery, duration of breastfeeding, and maternal education.

For fever, almost all (86%) the increase in inequality was accounted for by the increase in the effect of determinants of fever, while the distribution of determinants mattered less. The determinants with the greatest total contribution to the increase in inequality of fever were wealth, maternal education, birth order and breastfeeding duration.

In the multilevel model, we found that the likelihood of a child being stunted or experiencing fever depends on the community in which they live.

Conclusions

To curb the increase in inequality of stunting and fever, policy may focus on improving levels of, and reducing inequality in, access to facility deliveries, maternal nutrition (which may be related to maternal weight and height), complementary feeding (for breastfed children), wealth, maternal education, and child care (related to birth order effects). Improving overall levels of these determinants contribute to the persistence of inequality if these determinants are unequally concentrated on the well off to begin with.

1. Introduction

Socioeconomic inequalities in childhood health have persisted, with children from poor households experiencing a disproportionately larger burden (1, 2). This also implies that they may bear a larger share of later life consequences of childhood ill-health. Apart from increasing under-5 mortality rates, childhood ill-health negatively affects cognitive abilities, education attainment, later life income, and adult health (3-6). This study focusses on two key measures of childhood ill-health, namely, stunting and fever.

Fever is a broad measure of ill-health which may signal a number of sicknesses, including malaria and bacterial as well as viral infections (7-9). In children under the age of 5 years, high fever may also lead to seizures, brain damage or death (10). Similarly, stunting is a useful measure of childhood nutrition and ill-health. It is characterized by children being shorter than well-nourished kids of the same age and it is a culmination of chronic malnutrition or exposure to other adverse shocks. The adverse shocks may include illness, both in-utero (11) and in early life (12).

Of particular concern is the fact that fever and stunting, either directly or indirectly, constitutes a large share of childhood morbidity and mortality in low and middle income countries (13). Perhaps more concerning is the persistence of inequalities in childhood ill-health (2). Such inequalities are undesirable for at least two reasons. First, since inequalities in childhood health are related to inequalities in determinants of health—such as parental socioeconomic status, and access to health care, clean water, improved housing, neighborhoods etc.—which are predominately beyond the control of the child, and sometimes even parents themselves, it is considered unfair for some children to suffer more health challenges than others as a result of being at the disadvantage in accessing these determinants. Inequalities in child health are mostly a result of unequal opportunities in accessing determinants of good health so that such inequalities may be considered unnatural, avoidable and illegitimate. Avoidable inequalities are therefore termed inequities.

Second, persistence in inequalities in childhood ill-health is a source of concern on grounds that disease (whether infectious or not) in some parts of the population may affect the whole population (14). On a national level, health inequality may affect economic growth (15). Moreover, by plunging already poor households into health spending and possible labor income losses, e.g., due to taking care of sick children, health inequalities may widen income inequalities. Widening income inequalities are not only bad in themselves but they may also be harmful to the health of everyone in society, irrespective of their socioeconomic status (16).

Against this backdrop, reducing inequality has been espoused as one of the goals of the post 2015 development agenda, the Sustainable Development Goals (SDG). This is noteworthy because inequality was not an explicit goal in the Millennium Development Goals (MDGs), the predecessor to the SDGs. Despite the fact that policy documents in the MDG period emphasized the importance of reducing inequality,

which they argued was key to achieving the MDGs on health (17-19), there have been concerns that MDGs were not appropriate goals to drive the inequality agenda because they mainly focused on improving average health with little or no attention on how unequally the gains are distributed (20). An evaluation of the progress in child health in the MDG period leading up to 2015 shows that, despite acceleration in global reductions in under-5 mortality and a steady increase in life saving interventions, substantial inequalities in these interventions and in child health within and across countries have persisted (21). Given the multi-country nature of this evaluation, it remains unclear as to what factors could have been driving the persistence in inequality despite the substantial increase in life saving interventions. We use data from Zambia to understand inequalities in childhood ill-health. Zambia presents a unique opportunity in assessing inequality because of the recently conducted 2014 Demographic and Health Survey (DHS), a rich nationally representative household survey.

Zambia experienced sharp increases in a number of child health interventions in the MDG period (22, 23). At the same time, child mortality substantially reduced (23, 24). After remaining stubbornly high in the 1992--2001 period, under-5 mortality rate sharply declined in the 2001—2014 period, from 168 per 1000 live births to 75 per 1000 live births. In the 1992—2001 period, it only reduced from 191 to 168 per 1000 live births (24). The incidence as well as prevalence of key childhood ill-health also declined (24). Although inequalities in child mortality and ill-health have been documented(25), it is unclear how these inequalities evolved and what factors could have been driving these changes in the period leading up to 2015.

This paper uses large nationally representative household survey data from the Zambia Demographic and Health Survey (DHS) collected in 2007 and 2014 to examine changes in inequality in stunting and fever. Three objectives are pursued. First, and as a starting point, it explores the determinants of (factors associated with) stunting as well as fever and examines whether the community in which the child lives affects their health. Second, the paper investigates the significance of socioeconomic inequality in stunting and fever in 2007 and 2014 as well as whether or not the magnitude of inequality changed in any significant way over this period. The concentration index (CI) is used to quantify socioeconomic inequality in these measures of childhood ill-health.

Third, this paper examines how determinants of fever and stunting may explain changes in socioeconomic inequality as captured by the change in the CI over the 2007--2014 period. The change in the CI for stunting/fever is decomposed into the relative contribution of each determinant, which is further broken down into two components: changes in the CI of each determinant and changes in the effect of each determinant on stunting/fever, measured as an elasticity. By adding up the percentage contributions of each of the two components, we are able to look at the change in the CI of stunting/fever that was accounted

for by changes in CI of determinants on one hand and changes in the effect (elasticity) of determinants on the other hand.

Our study directly relates to studies from Vietnam that attempted to decompose the change in the concentration index of height for age. These studies found that rising inequality in height for age between 1990 and 2010 were mainly accounted for by both the increase in inequality in wealth and its elasticity (26, 27). The challenge with these studies is that they used data that does not contain a rich set of health variables. These missing variables may explain a significant portion of changes in inequality and may also confound the relationship between wealth and height for age. Other than height for age (or stunting), we are not aware of any study that has attempted to decompose changes in inequality of fever, as we do in this study.

Our study also relates to a rich literature that decomposes inequality in other dimensions of child health. As opposed to decomposing changes in inequality, such decompositions are only able to decompose a single concentration index and are thus not able to explain or quantify the sources of the observed change in two concentration indices that differ in time or space. Most studies that conduct decompositions over time have centered on explaining changes in average health and not changes in inequality (28).

This study also contributes to the literature that explores the effect of key determinants such as maternal education, wealth, maternal nutrition and other key covariates, on child health. Despite the fact that our estimates cannot be viewed as causal due to the cross sectional nature of our data, the rich set of covariates enables us to gain useful insight into the drivers of childhood ill-health.

In this paper, we say inequality to mean socioeconomic inequality in health as opposed to total inequality in health.

2. Data

Data were obtained from the 2007 and the 2014 Demographic and Health Survey (DHS). For children under the age of 5 years, our final dataset consisted of 5,616 observations in 2007 and 12,714 in 2014. The large difference in the number of observations between the two periods was due to the fact that the sample size for the 2014 DHS was more than doubled in order to provide reliable estimates for rural and urban areas within provinces (24).

The DHS uses a two stage sampling design where in the first stage, enumeration areas (clusters) are selected with probability proportional to size. The second stage selects households. In each household, three questionnaires are administered to eligible members by trained enumerators. The three questionnaires are the household's, woman's and man's questionnaires (24).

Child health information is captured in both the household's and woman's questionnaire. Since, we are interested in children under the age of 5 years, only women who had given birth within the five-year period preceding the relevant survey year were included. Using mother's identification variable, we merged the household and women data files. Children with missing mother identification variable either due to the mother being absent during the survey or due to incomplete interview were not included in the analysis.

Fever was measured by asking the mother whether her child had any fever within the two weeks preceding the survey. On the other hand, stunting was defined as having a height for age z-score of less than 2 standard deviations of the reference population using the WHO 2006 growth standards. Anthropometric measures (height and weight) were measured by the interviewer during each survey. Using the *zscore06* package in Stata (29), we computed height for age (HAZ) for each child. Consistent with the DHS methodology, HAZ was set to missing if height, age, or sex was either missing or out of range. All HAZ scores less than or greater than 6 standard deviation where regarded as out of range and dropped from the analysis.

The wealth index was calculated using principal component analysis and is provided together with the DHS data. Observations were ranked using the raw wealth index for purposes of computing concentration indices. In the decomposition analysis however, we grouped observations into quartiles.

3. Methods

3.1. Determinants of Stunting and Fever

For each survey year and each outcome, we fit a two level random intercept (multilevel) regression model. The first level is for the individual (child) while the second level is the community (enumeration area or cluster) where the child lives. The model takes the form:

$$y_{ijt} = \alpha_{jt} + \beta' x_{ijt} + \varepsilon_{ijt} \ldots\ldots\ldots\ldots 1$$

$$\alpha_{jt} = \delta_j + \mu_{jt} \ldots\ldots\ldots\ldots\ldots\ldots\ldots 2$$

where y_{ijt} is a binary variable equal to one if the outcome (fever or stunting) for child i residing in community j in year t is true. α_{jt}, is the random effect for community j in year t, with δ_j being the time average random effect for community j. x_{ijt} is a vector of determinants of y_{ijt} while β is a vector of regression coefficients which show the effect of x_{ijt} on y_{ijt}. The variable ε_{ijt} represents all other individual level determinants of y_{it} that we are not able to observe. It is normally distributed with mean zero and variance, $\sigma^2_{\varepsilon_{ijt}}$. Similarly, μ_{jt} represents all other community level unobservable determinants of child i's outcome. It has mean zero and variance, $\sigma^2_{\mu_{jt}}$. If variation at the community level , $\sigma^2_{\mu_{jt}}$, is sufficiently

small—approaching zero—then multilevel modelling is not necessary. We test the hypothesis that community level factors are not important determinants of childhood ill-health by assessing the size and significance of the intra-cluster correlation (ICC). The ICC is given as:

$$ICC = \frac{\sigma^2_{\mu_{jt}}}{\sigma^2_{\mu_{jt}} + \sigma^2_{\varepsilon_{ijt}}} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots 3$$

This paper does not aim to conduct a full multilevel analysis. Our only interest is to see whether or not, broadly viewed, the community in which a child lives matters for their health. As such, no covariates are included at the second level. We are only interested in the ICC and the coefficients in β.

The above regression model can be estimated using multilevel logistic regression since y_{ijt} is binary. Our interest is to also use the coefficients in β in the decomposition of the concentration index. However, since logistic regression is nonlinear while the decomposition of the concentration index requires linearity, we can either compute partial effects (probabilities) from the log odds, β, or use the log odds themselves in the decomposition.

Partial effects have the advantage of being easily understood. However, generating them from the vector β in multilevel logistic regression is complicated. Since we are interested in partial effects, and for ease of interpretation as well as computation simplicity, we used the multilevel linear regression which yields direct estimates of partial effects. Linear regression as a method of modelling binary variables, formally termed linear probability models (LPM), has seen widespread use in the literature lately and yields partial effects that are not different from probit or logistic regression partial effects (30-32). It has been shown that if interest is not in prediction but simply the coefficients vector, β, then the LPM is very appropriate (33).

3.2. Inequality in stunting and fever

We use the concentration index to quantify the extent of socio-economic inequality in the prevalence of stunting and incidence of fever in 2007 and 2014. The concentration index summarizes the extent to which good or bad health is dependent on income or wealth and it may be explained using the concentration curve concept. The concentration curve plots the cumulative share of health (on the y-axis) against the cumulative proportion of the population, ranked by wealth, from poor to richest (on the x-axis). For example, the concentration curve may show the cumulative percentage of stunting accruing to the poorest 25% of the population. To be complete, suppose that we want to look at inequality in ill-health. If the concentration curve lies on the 45-degree line, then the cumulative share of ill-health is equally shared

between the rich and the poor and there is no socioeconomic inequality in health. However, if the concentration curve lies on the left of the 45-degree line, then the poor carry a disproportionately high share of ill-health.

The standard concentration index is twice the area between the concentration curve and the 45-degree line and in any given year, *t*, it can be written as:

$$CI_{y_t} = \frac{2}{N\bar{y}_t}\sum_{i=1}^{N} y_{ijt}\, R_{ijt} - 1 \quad \text{.....4}$$

where $\bar{y}_t$ is the average rate of fever or stunting in year *t*. R_{ijt} is the rank of child *i*'s household in the wealth distribution, in our case measured by the wealth index from principal component analysis. The concentration index ranges from -1 to 1. It is zero if there is no socioeconomic inequality in health, -1 if all the ill-health is borne by the poor, and +1 if the richest have all the ill-health. It has been shown however that the concentration index may not be bounded between -1 and +1 if the health variable is binary (34), as it is in our case. This may lead to misleading conclusions. In particular, the bounds of the concentration index for a binary variable depend on average health and this can cause problems if one is comparing inequalities for two different areas or time periods that have substantially different average levels. This is important in our case since we compare inequality between 2007 and 2014.

Two alternative normalizations of the standard CI have been proposed by Wagstaff (34) and Erreygers (35). The standard CI is a measure of relative inequality, which is also the emphasis of the Wagstaff normalization. On the other hand, the Erreygers normalization is an absolute measure. It has been shown that neither of the two normalizations is superior to the other but each of them embodies different value judgements (36). We used the Wagstaff normalization in this paper. The normalization involves dividing the standard concentration index in Equation 4 by $(1 - \bar{y}_t)$ which give:

$$CI_{y_t} = \frac{\frac{2}{N\bar{y}_t}\sum_{i=1}^{N} y_{ijt}R_{ijt} - 1}{(1-\bar{y}_t)} \quad \text{.....5}$$

For each outcome, we computed this index in 2007 and 2014 to assess the extent of inequality in each year.

3.3. Changes in overall concentration Index

For each outcome, *y*, we computed the change in the concentration index as follows;

$$\Delta CI_y = CI_{y_{2014}} - CI_{y_{2007}} \quad \text{.....6}$$

The computation of the normalized CI based on Equation 5 and the change in the index as specified in Equation 6 involves a four stage computation process, which raises the issue of how to appropriately compute confidence intervals. In estimating the normalized CI for each year, the first stage involves the computation of the mean of the outcome and weighted fractional wealth rank for each year. In the second

stage, these estimates are combined to estimate the standard CI. The third stage involves dividing the standard CI by $(1-\bar{y}_t)$ to obtain the normalized CI. The change in the concentration index adds a fourth step to these computations; subtracting the 2007 normalized CI from that of 2014.

Our challenge is that since each estimate in these stages is computed from survey data, it has uncertainties which have to be taken into account when computing standard errors. Using analytical standard errors (from the last stage only) would make confidence intervals appear narrower than they actually are. To guard against this problem, we employ a bootstrap procedure with 1,000 replications. This involves repeating the above four step procedure 1,000 times, each time collecting the estimates, and then using these estimates to compute confidence intervals—which are then called bootstrap confidences intervals.

3.4. Decomposing changes in the concentration index

To decompose the changes in the overall concentration index, we make use of the estimated partial effects of determinants of fever/stunting, β, from Equation 1. The concentration index for outcome y in year t can then be written as a sum of the weighted concentration indices for all the determinants of y plus the generalized concentration index for the error term:

$$CI_{y_t} = \sum_k \left(\frac{\hat{\beta}_{kt}\bar{x}_{kt}}{\bar{y}_t}\right) CI_{kt} + \frac{GC_{\varepsilon t}}{\bar{y}_t} \quad \text{.....7}$$

where CI_{kt} is the concentration index for determinant k at time t computed as in Equation 5, that is, y_{ijt} in Equation 5 is replaced with x_{ijt} to get CI_{kt}. The weight, $\left(\frac{\hat{\beta}_{kt}\bar{x}_{kt}}{\bar{y}_t}\right)$ is the elasticity of the k^{th} variable with respect to the health variable y_{ijt} at time t and $GC_{\varepsilon t}$ is the generalized concentration index for the error term. $GC_{\varepsilon t}$ is obtained by multiplying the concentration index for the error term by the mean of the outcome, $\bar{y}_t$. Thus, $\frac{GC_{\varepsilon t}}{\bar{y}_t}$, is the concentration index for the error term. At any given time, t, Equation 7 says that the concentration index of y_t can be written as a weighted sum of the concentration indices of the K determinants plus the concentration index of the unobserved determinants of y_t. The weight for each concentration index of the determinant, CI_{kt}, is the elasticity of y_t with respect to that determinant (note that the elasticity is a nonlinear combination of $\hat{\beta}_{kt}$, $\bar{x}_{kt}$ and $\bar{y}_t$).

Equation 7 is the most commonly used method of decomposing inequalities in child health. Clearly this decomposition only allows one to examine the relative contribution of various determinants in explaining inequality at any given time, but it does not allow one to see which determinants are driving changes in inequality at any two given periods. To examining the drivers of changes in the childhood ill-health inequality specified in Equation 6 we apply the Oaxaca decomposition to Equation 7 (26). This leads to the following:

$$\Delta CI_y = \sum_k \eta_{k2014}(CI_{k2014} - CI_{k2007}) + \sum_k CI_{k2007}(\eta_{k2014} - \eta_{k2007}) + \Delta\left(\frac{GC_{\varepsilon t}}{\bar{y}_t}\right) \ldots\ldots\ldots\ldots\ldots 8$$

where η_{kt} is the elasticity of *y* with respect to determinant *k* in year *t*. Since $\eta_{kt} = \left(\frac{\hat{\beta}_{kt}\bar{x}_{kt}}{\bar{y}_t}\right)$, the elasticity of determinant k, η_{kt}, can change due to changes in any of its component, namely, $\bar{y}_t$, $\hat{\beta}_{kt}$, and $\bar{x}_{kt}$.

Equation 8 says that changes in the concentration index of health outcome *y* can be written as a sum of three components, namely, the weighted sum of the changes in the inequality of the *K* determinants, the weighted sum of the changes in the elasticities of *y* with respect to the *K* determinants, and the change in inequality of unobservable determinants. The change in inequality of each determinant is weighted by the elasticity of *y* with respect to this determinant in 2014 while the change in elasticity is weighted by the inequality of the determinant in 2007.

In other words, apart from the contribution of unexplained factors, $\Delta\left(\frac{GC_{\varepsilon t}}{\bar{y}_t}\right)$,the contribution of the k^{th} determinant to the change in inequality in *y*, ΔCI_y, can be brought about by the change in the concentration index of the k^{th} determinant, $(CI_{k2014} - CI_{k2007})$, or the change in it's the elasticity ,$(\eta_{k2014} - \eta_{k2007})$, or both. An increase in the concentration index of the k^{th} determinant in 2014 increases its contribution to inequality. On the other hand, the increase in its elasticity in 2014—resulting from a change in $\bar{y}_t$, $\hat{\beta}_{kt}$, $\bar{x}_{kt}$ or any other combination of these--can also contribute to the increase in inequality of childhood ill-health. For example, consider a case where the k^{th} determinant is concentrated on the well-off ($CI_{kt} > 0$) and it has a protective effect ($\hat{\beta}_{kt}$ is negative). In this case, a reduction in the prevalence of *y*, the mean $\bar{y}_t$, will increase inequality in *y*. Similarly, an increase in the mean of the k^{th} determinant,$\bar{x}_{kt}$will increase inequality. Holding $\bar{y}_t$ and $\bar{x}_{kt}$ constant, an increase in $\hat{\beta}_{kt}$ will also increase inequality.

4. Results

4.1. Descriptive statistics

The characteristics of children, mothers and households changed between the years 2007 and 2014. There was a substantial and significant increase in the proportion of children being delivered at a health facility in 2014 (Table 1). Birthweight was slightly lower in 2014 but the average duration of breastfeeding remained the same in both periods.

In 2014, mothers' education levels generally improved with significantly more mothers having secondary or higher education. Mothers of children under the age of five were also slightly larger in size -in term of height and weight- in 2014 and were also slightly older.

Living conditions also changed. The proportion of children coming from rural households was significantly lower in 2014. Access to improved sources of water increased substantially as did the proportion with improved toilets, although this increase was not as substantial.

There was no practically significant difference in household size and the number of under-5 children in the two periods.

Table 1: Descriptive Statistics

Variable	Mean		P-Value (Differences)
	2007 (N=5,616)	2014 (N=12,714)	H_0: $Mean_{2007}=Mean_{2014}$
Child's Characteristics			
Delivered at facility (%)	46.5	67.7	0.00
Birthweight (grams)	3238.1	3186.9	0.00
Childs age (months)	27.6	28.8	0.00
Duration of Breastfeeding (Months)	15.9	16.0	0.69
Birth Order	3.8	3.8	0.35
Mothers' Characteristics			
No Education (%)	13.7	11.1	0.00
Primary Education (%)	63.7	56.3	0.00
Secondary Education (%)	20.4	29.0	0.00
Higher Education (%)	2.2	3.6	0.00
Height (cm)	157.3	157.6	0.020
Weight (kg)	55.5	56.5	0.00
Age (years)	28.6	28.9	0.00
Employed (%)	59.4	59.1	0.78
Households' Characteristics			
Rural (%)	71.5	66.3	0.00
Improved Water Source (%)	35.7	59.5	0.00
Improved Toilet (%)	17.7	22.5	0.00
Household Size	6.2	6.5	0.00
Number of Children below 5 years	1.99	1.95	0.02

4.2. Regression Results

Clustering within communities

Table 2 shows two level random intercept models for stunting and fever by survey year. The intra cluster correlation for both stunting and fever are significantly different from zero implying that there is significant clustering of both stunting and fever. However, this clustering is higher for fever than it is for stunting.

Factors associated with stunting

In both years, lower height and lower weight of the mother was associated with a higher likelihood of stunting, with height exhibiting a particularly strong relationship. High birthweight was also associated

with lower likelihood of stunting in both years, while longer duration of breastfeeding and child being male were associated with higher likelihood of stunting.

Table 2: Effect of different factors on the probability of stunting and fever by year

	Stunting		Fever	
Variable	2007	2014	2007	2014
Wealth Quartile				
Quartile 1 (Poorest)	Base	Base		
Quartile 2	0.042 ((-0.023)— (0.106))	-0.032 ((-0.066)— (0.002))*	0.009 ((-0.020)— (0.039))	0.003 ((-0.019)— (0.024))
Quartile 3	0.031((-0.035)— (0.098))	-0.016 ((-0.053)— (0.021))	0.006((-0.028)— (0.040))	-0.018 ((-0.043)— (0.006))
Quartile 4 (Least poor)	-0.030((-0.114)— (0.053))	-0.042 ((-0.087)— (0.004))*	0.012((-0.038)— (0.063))	-0.034 ((-0.067)— (-0.002))**
Maternal Education				
No Education	Base	Base	Base	Base
Primary	0.008((-0.064)— (0.081))	-0.000 ((-0.040)— (0.040))	0.014((-0.018)— (0.045))	-0.028 ((-0.052)— (-0.003))**
Secondary	-0.016((-0.096)— (0.064))	-0.014 ((-0.058)— (0.031))	-0.002((-0.042)— (0.037))	-0.027 ((-0.057)— (0.002)) *
Higher	-0.038((-0.162)— (0.085))	-0.069 ((-0.138)— (-0.001))**	-0.042((-0.125)— (0.041))	-0.057 ((-0.109)— (-0.005))***
Other Maternal characteristics				
Age	-0.001((-0.007)— (0.004))	-0.004((-0.007)— (-0.001))***	-0.001((-0.004)— (-0.002))	-0.004((-0.006)— (-0.002))***
Height	-0.009((-0.012)— (-0.006))***	-0.010 ((-0.012)— (-0.008))***		
Weight	-0.002((-0.004)— (-0.000))**	-0.002((-0.003)— (-0.001))***		
Employed	-0.020((-0.058)— (0.019))	0.017((-0.006)— (0.039))	0.045((-0.023)— (0.067))***	0.036((0.020)— (0.053))***
Household Characteristics				
Rural	-0.001((-0.062)— (0.061))	-0.026((-0.057)— (0.005))	-0.006((-0.049)— (0.037))	-0.007((-0.033)— (0.020))
Improved water source	-0.000((-0.047)— (0.046))	-0.007((-0.033)— (0.018))	-0.009((-0.037)— (0.019))	-0.015((-0.033)— (0.003))
Improved toilet	-0.041((-0.092)— (0.009))	0.002((-0.025)— (0.028))	-0.009((-0.041)— (0.023))	-0.000((-0.020)— (0.020))
Household Size	-0.008((-0.017)— (0.002))	-0.006((-0.011)— (-0.001))**	0.006((0.001)— (0.012))**	0.007((0.003)— (0.011))***
Number of Children below 5 years	0.035((0.009)— (0.061))***	0.012((-0.003)— (0.026))	-0.020((-0.035)— (-0.006))***	-0.032((-0.042)— (-0.022))***
Child Characteristics				
Born at Facility	0.022((-0.041)— (0.085))	-0.016((-0.060)— (0.029))	-0.014((-0.037)— (0.010))	0.026((0.008)— (0.044))**
Birth Weight (kg)	-0.088((-0.117)— (-0.058))***	-0.095((-0.113)— (-0.077))***		
Male	0.070((0.033)— (0.106))***	0.058((0.037)— (0.079))***	0.022((0.002)— (0.042))**	-0.004((-0.018)— (0.012))
Age	-0.000((-0.002)— (0.001))	-0.001((-0.002)— (-0.000))**	-0.003((-0.004)— (-0.002))***	-0.002((-0.003)— (-0.001))***
Duration of breastfeeding	0.012((0.009)— (0.015))***	0.012((0.010)— (0.013))***	0.005((0.003)— (0.007))***	0.006((0.005)— (0.007))***
Birth order	-0.002((-0.018)— (0.014))	0.016((-0.006)— (0.025))***	-0.005((-0.013)— (0.004))	0.006((-0.001)— (0.012))
	Statistics			
Intra Cluster Correlation	0.025 ((0 .010)—(0.056))***	0.027((0 .017)—(0.042))***	0.047((0 .034)—(0.066))***	0.060((0 .049)—(0.073))***

Wealth and higher education level of the mother were associated with lower likelihood of stunting in 2014 but not in 2007. Similarly, in 2014, children of older mothers were less likely to be stunted than those with young mothers while children in higher birth order were more likely to be stunted.

Factors associated with fever

In both years, longer duration of breastfeeding was associated with higher likelihood of fever. A child whose mother was employed either in the formal or agricultural sector was more likely to experience fever compared to one whose mother was unemployed. Children from households that were larger were also more likely to have fever. Having a large number of children under 5 years in the household was associated with a lower likelihood of having fever. The likelihood of having fever is also lower the older the child.

As is the case with stunting, wealth and education were significantly associated with fever in 2014, but did not appear as important in 2007. In particular; any form of mothers' education was associated with lower likelihood of fever.

4.3. Socioeconomic Inequality in Childhood ill-health

Zambia had significant socioeconomic inequalities in stunting in both 2007 and 2014 (Table 3). The negative sign of the concentration indices indicates that children from poorer households carried a disproportionally higher burden of stunting than their relatively better-off counterparts. In spite of the reduction in stunting prevalence from 45.6% in 2007 to 40% in 2014, the levels of inequality as measured by the Wagstaff CI significantly increased by about 45%, from -0.093 in 2007 to -0.135 in 2014. The increase in the concentration index implies that stunting was reduced less among the poor. In fact, a tabulation of stunting levels by wealth quartiles (not reported) shows that the poorest quartile did not register any change in stunting.

Table 3: Mean levels and socioeconomic inequality in stunting and fever in 2007 and 2014

	2007	**2014**	**H_0: $Y_{2007}=Y_{2014}$**
	Estimate(95% Bootstrap CI)	Estimate(95% Bootstrap CI)	Bootstrap P-Value
Prevalence of Stunting			
Mean	0.456 (0.442—0.471)	0.400 (0.389—0.410)	0.000
Concentration Index	-0.093 ((-0.128)—(-0.058))	-0.135 ((-0.160)—(-0.109))	-0.041(0.051)
Incidence of Fever			
Mean	0.184 (0.173—0.195)	0.216 (0.208—0.225)	0.000
Concentration Index	-0.015 ((-0.057)—(0.027))	-0.064 ((-0.092)—(-0.036))	-0.049 (0.055)

In 2007, there is no evidence of inequality in fever incidence (Table 3). Inequalities however, emerged in 2014, with a concentration index of -0.064. The incidence of fever also increased slightly from 18.4% to 21.6%.

4.4. Decomposition of Changes in Concentration Index

Explaining changes in the inequality in stunting

The concentration index of stunting increased by -0.041 (became more pro-poor-- i.e., stunting was reduced less among the poor so that inequalities increased) between 2007 and 2014 (Table 3). This increase in inequality of stunting was accounted for by both the increase in the CI of determinants (42.5%) and the increase in the effect of determinants (35%), measured as elasticities (Table 4). The rest of the increase (22.5%) was due to unexplained factors.

Table 4: Decomposition of the change in inequality in stunting and fever

	Stunting				Fever			
	Weighted Δ in CI	Weighted Δ in elasticity	Total Δ	Total % Δ	Weighted Δ in CI	Weighted Δ in elasticity	Total Δ	Total % Δ
Wealth Quartile								
Quartile 1 (Poorest)								
Quartile 2	0.001	0.009	0.010	-25	-0.000	0.002	0.001	-3
Quartile 3	0.001	-0.007	-0.006	16	0.001	-0.007	-0.006	11
Quartile 4 (Least poor)	0.001	-0.007	-0.007	16	0.001	-0.044	-0.043	88
Maternal Education								
No Education								
Primary	0.000	0.001	0.001	-3	0.006	0.009	0.015	-30
Secondary	0.000	-0.001	-0.000	0	0.001	-0.013	-0.011	23
Higher	0.000	-0.004	-0.003	9	0.000	-0.005	-0.004	9
Other Maternal characteristics								
Age	-0.003	0.002	-0.000	1	-0.005	0.004	-0.001	2
Height	-0.009	-0.001	-0.010	25				
Weight	-0.004	-0.000	-0.005	12				
Employed	0.000	-0.004	-0.004	9	0.001	0.002	0.002	-5
Household Characteristics								
Rural	0.001	0.011	0.012	-30	0.000	0.000	0.001	-2
Improved water source	0.003	-0.005	-0.002	6	0.009	-0.012	-0.003	5
Improved toilet	-0.000	0.010	0.009	-23	0.000	0.003	0.003	-7
Household Size	-0.001	0.000	-0.001	2	0.002	0.000	0.002	5
Number of Children below 5 years	0.000	0.004	0.004	-10	-0.001	0.003	0.002	-5

Child Characteristics								
Born at Facility	0.003	-0.014	-0.010	26	-0.010	0.030	0.019	-39
Birth Weight (kg)	-0.001	0.000	-0.001	3				
Male	-0.001	-0.000	-0.001	2	0.000	-0.001	-0.001	1
Age	-0.000	-0.000	-0.000	1	-0.001	0.000	-0.001	1
Duration of breastfeeding	-0.006	0.000	-0.005	13	-0.006	-0.002	-0.008	16
Birth order	-0.001	-0.010	-0.011	27	-0.000	-0.011	-0.012	24
Residuals			-0.009	24			-0.007	14
Total	-0.017	-0.014	-0.041	100	-0.000	-0.042	-0.049	100
Percent of total Δ*	42.5	35		100*	0	86		100*

TABLE NOTES: Table 4 shows the decomposition of the change in CI according to Equation 8 for stunting (first 4 columns) and fever (next 4 columns). The total change is given in column 3 and 7 of the last but one row. The variables that contributed positively to this increase have a negative quantity in column 3 and 7 (negative because they made the CI more negative—increased concentration of ill-health on the poor). This translates to a positive percentage change in contribution to inequality (Column 4 and 8).

*This adds column 1 and 2, the difference is due to residuals.

The determinants that contributed most to the increase in inequality of stunting were mother's height and weight (37%), being in the two wealthiest quartiles (32%), birth order (27%), facility delivery (26%), duration of breastfeeding (13%), and higher level of maternal education (9%). Other factors worked to reduce inequality and hence have negative percentage contribution to the increase in inequality. But how did the change in CI and effect of each of these determinants contribute to the increase in inequality of stunting? Table 4 shows that the CI for height and weight increased (became more pro-rich—heights and weights increased more for the rich) while at the same time the effect of these determinants on stunting increased. These two mechanisms reinforced each other to drive inequality in stunting up, with the increase in the CI having a particularly larger contribution. On the other hand, since the CI of wealth itself reduced, the contribution of wealth to the increase in inequality (in the top 2 quartiles) was solely due to the increase in the effect of wealth. The change in the CI of birth order and the change in the effect of birth order on stunting reinforced each other to drive inequality up. In particular, the increased effect of birth order on stunting was both a result of higher birth order becoming more significantly associated with increased likelihood of stunting in 2014 (Table 2) and a reduction in the prevalence of stunting (Table 3). At the same time, birth order became more concentrated among the poor implying that the poor bore a disproportionately larger share of the risk arising from higher birth order.

Since the CI of facility deliveries itself reduced (became less pro-rich—facility deliveries increased more among the poor), the contribution of facility deliveries to the increase in inequality of stunting was almost entirely driven by the increase in its effect on stunting. Despite the reduction in the CI, however, inequality in facility deliveries remained pro-rich (the rich still had higher access to facility deliveries). Hence the increase in the protective effect of facility deliveries disproportionately benefited the better off. The same

can be said about maternal education; the effect strengthened but this benefit accrued more to the better off since they had a disproportionately larger share of higher education. The contribution of duration of breastfeeding to the increase in inequality was entirely due to the inequality effect; longer periods of breastfeeding becoming more concentrated on the poor.

Explaining changes in the inequality in fever

While both changes in the CI (inequality) of determinants as well as the effects (elasticity) of determinants were important in explaining increasing inequality of stunting, almost all (86%) the increase in the inequality of fever incidence was accounted for the change in the effects of determinants (Table 4). The changes in the CI of determinants, overall, did not explain any increase in the inequality of fever, implying that the rest of the increase--14%--was accounted for by unexplained/unobserved determinants. The key contributors to the increase in inequality in fever incidence were wealth (99%), mother's education (32%), birth order (24%), and duration of breastfeeding (16%). Note that the overall contribution of all determinants add up to 100% because other determinants worked to reduce inequality (have negative percentage contribution)

As indicated earlier, the CI for wealth reduced slightly although it remained pro-rich. However, there was a substantial strengthening of the effect of wealth on fever in 2014, and due to the highly pro-rich distribution of wealth, most of this benefit accrued to the better off. This drove inequality in fever up, as did maternal secondary or higher education.

Even if almost all the increase in inequality of fever was accounted for by the increase in the effect of determinants, some determinants' contribution were both due to the change in their effect and their concentration indices. This can be said of birth order whose contribution to the increase in inequality of fever was due to the two mechanisms reinforcing each other— higher birth order becoming more concentrated on the poor and a strengthening of effect of birth order on fever.

5. Discussion

We investigated determinants of, and socioeconomic inequality in, stunting and fever in Zambia between 2007 and 2014, a period when child health interventions were rapidly scaled up to meet the 2015 MDG target on child health. We find that although stunting prevalence reduced, inequality increased. On the other hand, fever incidence did not fall but inequality still increased. The increase in inequality of stunting and fever implies that the rapid scale up of child health interventions may not have been successful in reducing childhood disease burden among the most vulnerable, suggesting the need for policy reform if the goal of reducing inequality, as captured by the Sustainable Development Goals (SDGs), is to be achieved.

We also find evidence of clustering for both stunting and fever implying that the likelihood of being stunted or having fever partly depends on the area in which the child lives, which is particularly apparent for fever. Elsewhere, fever has also been shown to exhibit substantial clustering (37-39), and in many African countries, it is highly associated with malaria and pneumonia (40). Clustering is a form of inequality and implies that some areas suffer higher burden of childhood ill-health than others.

Our study included a rich set of determinants that potentially explain the likelihood of a child getting fever or being stunted. We document a very strong association between maternal size (height and weight) and stunting. The association between maternal height and weight on one hand and stunting on the other hand may be due to genetic factors or maternal nutritional deficiencies—showing up as low maternal weight and short maternal height. Maternal nutritional deficiencies may lead to in-utero growth restriction, (41), so that children whose growth was restricted end up being stunted. However, while low maternal weight may be directly related to nutrition during pregnancy which directly impacts on fetal growth, maternal height is related to nutrition during the mother's own childhood and only affects fetal growth indirectly through, for example, smaller sizes of reproductive organ, reduced protein and energy stores, and limited room for child development in utero (42-44). Our findings are consistent with a multi-county randomized trial that included Zambia (45). Since there is a strong correlation between maternal height, weight and social economic status, our results imply that previous studies investigating inequality in stunting that did not control for maternal height and weight may have overestimated the effect of socioeconomic variables such as wealth and education.

In line with the strong and consistent correlation, the decomposition analysis showed that inequality of maternal height and weight was the biggest driver of the increase in inequality in stunting over the period 2007--2014. The increase in inequality was mainly due to the fact that more advantageous heights and weights became more concentrated on wealthier mothers. It is therefore important to reduce inequality in maternal nutrition, both early in life and during pregnancy, to halt increases in inequality in stunting.

Another interesting finding relates to birth order. We found that higher birth order was a risk factor for stunting and fever. These findings are consistent with a number of studies that have documented a negative association between higher birth order and child health (46-49), education attainment (50-54) as well as cognitive abilities (55). Debate on the exact mechanism through which higher birth order is a risk factor for most outcomes seems to be polarized with others indicating that the cause is biological and others indicating that it is confounded by family size, a variable we control for in our set-up. Yet others have pointed to the social interaction mechanism where children born later receive less favorable social interactions. Consensus seems to have emerged that the social interaction mechanism is the cause of the observed association (55, 56). We have documented that birth order contributed to the increase in inequality of both stunting and fever as higher birth orders became even more concentrated among the poor. If the

social interaction hypothesis is true, it may be beneficial to use routine health programs to emphasize the importance of child care for children of higher birth orders.

We also document a consistent correlation between duration of breastfeeding and childhood ill-health where possible confounders, including wealth, are adjusted for. In results not reported, this correlation was generally maintained in all households after stratifying by wealth quartile. There is mixed evidence on the effect of breastfeeding duration on child health. A number of studies find a positive correlation between duration of breastfeeding and poor growth (57-62) while others do not. It is generally held that this positive correlation is due to two possible mechanisms. First is the case of reverse causality were children who are in poor health to begin with are breastfed for longer. We cannot rule out this possibility since we are using cross sectional data. Note however that the relationship between breastfeeding duration and childhood ill-health in our set up is not likely to be driven by differences in wealth, maternal education, maternal nutrition, maternal age, child's birthweight, etcetera, because we control for these possible confounders in the regression analysis. Second, there is possibility that sufficient complementary food is not provided to meet energy and nutritional demands of the child (63). The insufficient feeding argument seems compelling given that longer breastfeeding duration, while possibly having other benefits, may not be helpful for child growth without sufficient complementary feeding (64-66). Proteins, necessary for growth, may become deficient if there is laxity in providing adequate complementary food and more emphasis is placed on breastfeeding (65). If we are comparing children of the same birthweight, age, sex and coming from equally wealthy households, then laxity in providing complementary feeding may explain the observed relationship.

Although longer breastfeeding duration was a risk in all wealth quartiles, poorer households breastfed longer than their well off counterparts. The concentration of longer breastfeeding duration on poorer households contributed to the increase in inequality of childhood illness. One should however, interpret these results with caution because breastfeeding in itself has been shown to have other beneficial effects (67). Perhaps what comes out from this finding is that programs should emphasize sufficient complementary finding even with longer breastfeeding to ensure sufficient protein and energy intake. It has been shown that even if children from low income countries, in general, start at the same average height for age as the reference population, there is rapid faltering of growth in the first 2 years of life (68, 69). This faltering may suggest inadequate quality and quantity of complementary foods. However, disease may also explain this poor growth given a complex interaction between malnutrition and disease.

Facility deliveries may be an important entry point for feeding counselling and support interventions. This assertion is supported by evidence from Uganda were being delivered at a facility was associated with better child feeding practices and nutritional status of children (70). We find that the increase in the effect of facility deliveries (in terms of elasticity) contributed to the increase in inequality of stunting (driven by

the increase in mean level of facility deliveries, reduction in mean stunting and improvement in the association between facility delivery and stunting). However, increases in facility deliveries mainly benefited the well off more because they were unequally concentrated on the well off to begin with. Thus, the increased effect also contributed to the increase in the inequality of stunting.

The question emerges, how can policy halt the increasing socioeconomic inequalities in stunting prevalence and fever incidence?

One possible option would be to reduce, and possibly eliminate, the effect of determinants that increase the risk of stunting and fever. This can be done in three ways, as can be seen when the elasticity formula is unpacked (we discussed this mechanism in the last part of the methods section). First, the average level of determinants that are risk factors for child health can be reduced. Second, the association (marginal effect) between (of) the determinant and (on) child health can be eliminated. For example, in the case of birth order, routine health education programs may emphasize the importance of giving as much attention to children of higher birth order as those of lower birth order. This may diminish the association and thus reduce inequality. Third, and lastly, the incidence or prevalence of the relevant childhood ill-health can be reduced.

The other option would be to reduce inequality (concentration indices) in (of) determinants such as wealth, education, maternal nutrition, etcetera, which are protective for child health. These determinants are unequally concentrated on the well off and reducing their inequality may be beyond the scope of health interventions. It is important to realize that health interventions can mainly affect the effect, in terms of elasticity, of these determinants on child health but not their distribution. For example, since more educated mothers are likely to understand and follow health instructions better, health interventions such as infant and child feeding, breastfeeding counselling, etc., may increase the association between maternal education and stunting but are not able to reduce inequality in education. The same can be said about wealth. Therefore, it is easy to see why, despite the rapid increase in child health intervention coverage (such as infant and child feeding counselling and support) and the accompanying increase in the effects of determinants, inequality in childhood-ill-health still increased. The dichotomy is that increasing the effect of determinants with protective effects, which is a good thing in general, worsens inequality if these determinants are disproportionately concentrated on the rich in the first place. This is an example of the classical equity-efficiency trade off. Despite the persistence of inequality, other studies have documented substantial improvements in other measures of child health, such as under-five mortality (24, 71). Our findings suggest that one of the reasons for the persistent inequalities is that health improvements disproportionately benefit the well off because the determinants of childhood good health such as access to health facilities, which health and other interventions seek to improve, are unequally concentrated on the well-off, to begin with.

Our study has limitations and due caution must be excised when interpreting the findings. It must be noted that no community level covariates were included at the second stage of the multilevel model. This raises a possibility of confounding if the omitted community level covariates are correlated with both childhood ill health (stunting or fever) and any included individual or household level variable. Moreover, being based on cross-sectional data, our results cannot be viewed as causal.

It is worth mentioning, however, that the data we have used, the Demographic and Health Survey (DHS) presents both limitations and strengths for inequality analyses. It is a limitation because DHS relies on wealth indices and does not contain finer measures of household living standards, such as consumption or income. Income or consumption has the advantage of being able to be objectively measured and compared across different places or surveys with less difficulty. Wealth indices may be problematic if one is comparing two populations (e.g rural with urban or population in 1970 with population in 2014) as the type of assets and their valuation may differ across populations and time. Although some methods on how to make wealth indices collected in two different populations or two different points of time have been proposed(72), they are, at best, imperfect. To reduce this comparability problem, analyses, and concentration indices were calculated separately for each year.

Using the DHS is also a strength of our study. This is because it contains a rich set of health variables. As an alternative, we could have followed studies from Vietnam and used data from the Living Conditions Measurement Surveys(LCMS)which have information on income, consumption and anthropometric measurements. However, the LCMS does not contain a rich set of health variables, as does the DHS, and this can potentially confound the relationship between key socioeconomic variables and child health. It also does not contain other child health outcomes such as fever.

6. Conclusion

Childhood ill-health has serious consequences. Apart from increasing under-5 mortality rates, it negatively affects cognitive abilities, education attainment, later life income, and adult health. However, children in low socioeconomic background bear a significantly larger share of childhood ill-health implying that they will continue to shoulder a larger share of these adverse consequences. This raises ethical issues. Why should children from poor backgrounds experience more ill-health when the determinants of ill-health are beyond their control, and to a large extent beyond the control of their parents. How can such inequalities be justified when they are hugely generated by inequality of opportunities to determinants of good health, such as education and health care? Against this backdrop, reducing inequality constitutes one of the most important development goals and is now part of the post 2015 development agenda, the Sustainable Development Goals (SDG), the successor to the Millennium Development Goals (MDG). To derive lessons for the post 2015 agenda of designing interventions that are effective in improving overall child health

and reducing inequality, it is important go beyond asking whether or not inequalities increased by undertaking an in-depth analysis of the forces that drive inequality.

We examined the determinants of stunting and fever using the 2007 and 2014 Zambia DHS data to explore the existence of socioeconomic inequalities in childhood ill-health indicators, and whether this inequality changed over the period. Most importantly, we quantified how changes in inequality of determinants and the changes in elasticity of stunting and fever with respect to their determinants could have contributed to the change in inequality between 2007 and 2014.

While the prevalence of stunting reduced substantially, inequality increased between 2007 and 2014. In fact, inequalities were worsened by the fact that the prevalence of stunting was reduced in all quartiles, except the poorest. This increase in inequality was largely a result of the increase in the inequality of factors that are associated with stunting. These factors include maternal height and weight, wealth, birth order, breastfeeding duration, facility deliveries and maternal education. Although the responsive of stunting to most of these factors, e.g. facility deliveries, increased in 2014, this benefit mostly accrued to the better off because the factors remained concentrated on the better off. As a consequence, the improved responsiveness of stunting to its determinants also contributed to the increase in inequality.

Regarding fever, almost all the increase in the inequality was account for by the increase in the responsiveness of the disease to the factors that determine it. By far the biggest driver of this change was wealth, then maternal education, birth order and breastfeeding duration.

The key message in this study is that halting the increase in the inequality in childhood ill-health depends heavily on reducing inequality in the factors that affect childhood ill-health while at the same time improving the impact (elasticity) of these factors using both health and non-health interventions. It is important to note that although improving impact of factors that affect child health is desirable, this in itself can be a source of increase in inequality if the factors whose impact are being improved are unequally concentrated on the better off to begin with.

Halting the increase in child health inequality is dependent on reducing inequality in the factors associated with child health. These include wealth, maternal education, appropriate feeding and weaning (related to breastfeeding duration effects), adequate care giving (related to birth order effects) and maternal nutrition (related to maternal height and weight). We believe that a more sustainable way of doing this is to ensure equality of opportunities in access to these factors among children from different socioeconomic backgrounds—who are future parents. This may call for policies that delink the dependence of child health on parental circumstances and the community they live. Specifically, group specific interventions aimed at the most vulnerable may need to be implemented along with population level interventions. For example, under the current social cash transfer scheme in Zambia, policy may aim at providing more cash benefits to the poorest households who have children under the age of five years. Poor households may also have

special educational needs. For example, children from poor backgrounds may have challenges learning and concentrating even when given access to school due to persistent hunger. Child school feeding programs may help in improving attendance among poor children and also reducing inequality in learning and concentration.

Moreover, despite the fact that all households require appropriate breastfeeding and weaning educational interventions--due to the observed association between duration of breastfeeding and child health--poorer households may require special interventions since they have disproportionately longer breastfeeding durations. In general, the propensity to breastfeed longer may be due to lack of appropriate food or knowledge on how to use existing traditional food stuffs. Thus, policy may focus on introducing and scaling up complementary feeding and nutritional programs among poor households. Additionally, since the community in which a child lives matters for child health, policy may also focus on improving living conditions in disadvantaged communities, e.g. sanitation facilities, water, child care centers, etc. The implementation of such group specific intervention may enhance equality of opportunity and halt the increase in child health inequality.

7. References

1. Quentin W, Abosede O, Aka J, Akweongo P, Dinard K, Ezeh A, et al. Inequalities in child mortality in ten major African cities. BMC Medicine. 2014;12(1):1-12.
2. Bredenkamp C, Buisman LR, Van de Poel E. Persistent inequalities in child undernutrition: evidence from 80 countries, from 1990 to today. International Journal of Epidemiology. 2014;43(4):1328-35.
3. Alderman H, Hoddinott J, Kinsey B. Long term consequences of early childhood malnutrition. Oxford Economic Papers. 2006;58(3):450-74.
4. Currie J, Rossin-Slater M. Early-Life Origins of Life-Cycle Well-Being: Research and Policy Implications. Journal of Policy Analysis and Management. 2015;34(1):208-42.
5. Almond D, Currie J. Killing Me Softly: The Fetal Origins Hypothesis. The journal of economic perspectives : a journal of the American Economic Association. 2011;25(3):153-72.
6. Berkman DS, Lescano AG, Gilman RH, Lopez SL, Black MM. Effects of stunting, diarrhoeal disease, and parasitic infection during infancy on cognition in late childhood: a follow-up study. Lancet. 2002;359(9306):564-71.
7. Kwiatkowski D. Tumour necrosis factor, fever and fatality in falciparum malaria. Immunology Letters. 1990;25(1):213-6.
8. Gonzalez Plaza JJ, Hulak N, Zhumadilov Z, Akilzhanova A. Fever as an important resource for infectious diseases research. Intractable & rare diseases research. 2016;5(2):97-102.
9. Luszczak M. Evaluation and management of infants and young children with fever. American family physician. 2001;64(7):1219-26.
10. Green R, Jeena P, Kotze S, Lewis H, Webb D, Wells M. Management of acute fever in children: guideline for community healthcare providers and pharmacists. SAMJ: South African Medical Journal. 2013;103(12):948-54.
11. Banerjee A, Duflo E, Postel-Vinay G, Watts T. Long-Run Health Impacts of Income Shocks: Wine and Phylloxera in Nineteenth-Century France. Review of Economics and Statistics. 2010;92(4):714-28.

12. Lee G, Yori P, Olortegui MP, Pan W, Caulfield L, Gilman RH, et al. Comparative effects of vivax malaria, fever and diarrhoea on child growth. International Journal of Epidemiology. 2012;41(2):531-9.
13. Murray CJL, Vos T, Lozano R, Naghavi M, Flaxman AD, Michaud C, et al. Disability-adjusted life years (DALYs) for 291 diseases and injuries in 21 regions, 1990–2010: a systematic analysis for the Global Burden of Disease Study 2010. The Lancet. 2012;380(9859):2197-223.
14. Woodward A, Kawachi I. Why reduce health inequalities? Journal of Epidemiology and Community Health. 2000;54(12):923-9.
15. Grimm M. Does inequality in health impede economic growth? Oxford Economic Papers. 2011.
16. Pickett KE, Wilkinson RG. Income inequality and health: A causal review. Social Science & Medicine. 2015;128:316-26.
17. Jong-wook L. Global health improvement and WHO: shaping the future. The Lancet. 2003;362(9401):2083-8.
18. WHO. The world health report 2003: shaping the future. Geneva: World Health Organization; 2003.
19. UN. Implementing the Millennium Development Goals: Health Inequality and the Role of Global Health Partnerships. Ney York: United Nations, Policy CfD; 2009 July 2009. Report No.
20. Kabeer N. Can the MDGs provide a pathway to social justice? The challenge of intersecting inequalities. United Nations Development Programme, New York: 2010.
21. Victora CG, Requejo JH, Barros AJD, Berman P, Bhutta Z, Boerma T, et al. Countdown to 2015: a decade of tracking progress for maternal, newborn, and child survival. The Lancet. 2016;387(10032):2049-59.
22. Colson KE, Dwyer-Lindgren L, Achoki T, Fullman N, Schneider M, Mulenga P, et al. Benchmarking health system performance across districts in Zambia: a systematic analysis of levels and trends in key maternal and child health interventions from 1990 to 2010. BMC Medicine. 2015;13(1):1-14.
23. Ng M, Colson KE, Fullman N, Dwyer-Lindgren L, Achoki T, Schneider MT, et al. Assessing the Contribution of Malaria Vector Control and Other Maternal and Child Health Interventions in Reducing All-Cause Under-Five Mortality in Zambia. The American Journal of Tropical Medicine and Hygiene. 2016.
24. Central Statistical Office (CSO) [Zambia], Ministry of Health (MOH) [Zambia], ICF International. Zambia Demographic and Health Survey 2013-14. Rockville, Maryland, USA: Central Statistical Office, Ministry of Health, and ICF International., 2014.
25. World Bank. Health equity and financial protection datasheet : Zambia. Health equity and financial protection datasheet. . Washington, DC World Bank Group., 2012 2012/08/01. Report No.: Contract No.: 72211.
26. Wagstaff A, van Doorslaer E, Watanabe N. On decomposing the causes of health sector inequalities with an application to malnutrition inequalities in Vietnam. Journal of Econometrics. 2003;112(1):207-23.
27. Kien VD, Lee H-Y, Nam Y-S, Oh J, Giang KB, Van Minh H. Trends in socioeconomic inequalities in child malnutrition in Vietnam: findings from the Multiple Indicator Cluster Surveys, 2000–2011. Global Health Action. 2016;9:10.3402/gha.v9.29263.
28. O'Donnell O, Nicolás ÁL, Van Doorslaer E. Growing richer and taller: Explaining change in the distribution of child nutritional status during Vietnam's economic boom. Journal of Development Economics. 2009;88(1):45-58.
29. Leroy J. zscore06: Stata command for the calculation of anthropometric z-scores using the 2006 WHO child growth standards. 2011.
30. De Neve J-W, Fink G, Subramanian SV, Moyo S, Bor J. Length of secondary schooling and risk of HIV infection in Botswana: evidence from a natural experiment. The Lancet Global Health. 2015;3(8):e470-e7.

31. Kröger H. The contribution of health selection to occupational status inequality in Germany – differences by gender and between the public and private sectors. Public Health. 2016;133:67-74.
32. Friedman J, Schady N. HOW MANY INFANTS LIKELY DIED IN AFRICA AS A RESULT OF THE 2008–2009 GLOBAL FINANCIAL CRISIS? Health Economics. 2013;22(5):611-22.
33. Wooldridge JM. Econometric analysis of cross section and panel data: MIT press; 2010.
34. Wagstaff A. The bounds of the concentration index when the variable of interest is binary, with an application to immunization inequality. Health Economics. 2005;14(4):429-32.
35. Erreygers G. Correcting the Concentration Index. Journal of Health Economics. 2009;28(2):504-15.
36. Kjellsson G, Gerdtham U-G. On correcting the concentration index for binary variables. Journal of Health Economics. 2013;32(3):659-70.
37. Sluydts V, Heng S, Coosemans M, Van Roey K, Gryseels C, Canier L, et al. Spatial clustering and risk factors of malaria infections in Ratanakiri Province, Cambodia. Malaria Journal. 2014;13(1):1-12.
38. Ngianga-Bakwin K, Madise N. The spatial epidemiology of childhood diseases in Malawi and Zambia. African Population Studies. 2004;19(Supplement B):199-226.
39. Kandala N-B, Magadi MA, Madise NJ. An investigation of district spatial variations of childhood diarrhoea and fever morbidity in Malawi. Social Science & Medicine. 2006;62(5):1138-52.
40. Yeboah-Antwi K, Pilingana P, Macleod WB, Semrau K, Siazeele K, Kalesha P, et al. Community Case Management of Fever Due to Malaria and Pneumonia in Children Under Five in Zambia: A Cluster Randomized Controlled Trial. PLoS Med. 2010;7(9):e1000340.
41. Katona P, Katona-Apte J. The Interaction between Nutrition and Infection. Clinical Infectious Diseases. 2008;46(10):1582-8.
42. Duggleby SL, Jackson AA. Relationship of maternal protein turnover and lean body mass during pregnancy and birth length. Clinical science (London, England : 1979). 2001;101(1):65-72.
43. Addo OY, Stein AD, Fall CH, Gigante DP, Guntupalli AM, Horta BL, et al. Maternal Height and Child Growth Patterns. The Journal of Pediatrics. 2013;163(2):549-54.e1.
44. Ibanez L, Potau N, Enriquez G, de Zegher F. Reduced uterine and ovarian size in adolescent girls born small for gestational age. Pediatric research. 2000;47(5):575-7.
45. Krebs NF, Mazariegos M, Chomba E, Sami N, Pasha O, Tshefu A, et al. Randomized controlled trial of meat compared with multimicronutrient-fortified cereal in infants and toddlers with high stunting rates in diverse settings. The American Journal of Clinical Nutrition. 2012;96(4):840-7.
46. Myrskylä M, Silventoinen K, Jelenkovic A, Tynelius P, Rasmussen F. The association between height and birth order: evidence from 652 518 Swedish men. Journal of Epidemiology and Community Health. 2013.
47. Rahman M. Association between order of birth and chronic malnutrition of children: a study of nationally representative Bangladeshi sample. Cadernos de saude publica. 2016;32(2):e00011215.
48. Elliott BA. Birth Order and health: major issues. Social science & medicine (1982). 1992;35(4):443-52.
49. Modin B. Birth order and mortality: a life-long follow-up of 14,200 boys and girls born in early 20th century Sweden. Social Science & Medicine. 2002;54(7):1051-64.
50. Black SE, Devereux PJ, Salvanes KG. The More the Merrier? The Effect of Family Size and Birth Order on Children's Education. The Quarterly Journal of Economics. 2005;120(2):669-700.
51. Härkönen J. Birth Order Effects on Educational Attainment and Educational Transitions in West Germany. European Sociological Review. 2014;30(2):166-79.
52. Tenikue M, Verheyden B. Birth Order and Schooling: Theory and Evidence from Twelve Sub-Saharan Countries. Journal of African Economies. 2010;19(4):459-95.
53. De Haan M, Plug E, Rosero J. Birth Order and Human Capital Development: Evidence from Ecuador. Journal of Human Resources. 2014;49(2):359-92.

54. Mogstad M, Wiswall M. Testing the quantity–quality model of fertility: Estimation using unrestricted family size models. Quantitative Economics. 2016;7(1):157-92.
55. Kristensen P, Bjerkedal T. Explaining the Relation Between Birth Order and Intelligence. Science. 2007;316(5832):1717-.
56. Rodgers JL, Cleveland HH, van den Oord E, Rowe DC. Resolving the debate over birth order, family size, and intelligence. American Psychologist. 2000;55(6):599-612.
57. Haschke F, van't Hof MA. Euro-Growth references for breast-fed boys and girls: influence of breast-feeding and solids on growth until 36 months of age. Euro-Growth Study Group. Journal of pediatric gastroenterology and nutrition. 2000;31 Suppl 1:S60-71.
58. Hediger ML, Overpeck MD, Ruan WJ, Troendle JF. Early infant feeding and growth status of US-born infants and children aged 4-71 mo: analyses from the third National Health and Nutrition Examination Survey, 1988-1994. Am J Clin Nutr. 2000;72(1):159-67.
59. Nielsen GA, Thomsen BL, Michaelsen KF. Influence of breastfeeding and complementary food on growth between 5 and 10 months. Acta Paediatr. 1998;87(9):911-7.
60. Spyrides MH, Struchiner CJ, Barbosa MT, Kac G. Effect of predominant breastfeeding duration on infant growth: a prospective study using nonlinear mixed effect models. Jornal de pediatria. 2008;84(3):237-43.
61. WHO. An evaluation of infant growth: the use and interpretation of anthropometry in infants. WHO Working Group on Infant Growth. Bull World Health Organ. 1995;73(2):165-74.
62. Kattula D, Sarkar R, Sivarathinaswamy P, Velusamy V, Venugopal S, Naumova EN, et al. The first 1000 days of life: prenatal and postnatal risk factors for morbidity and growth in a birth cohort in southern India. BMJ open. 2014;4(7):e005404.
63. Griffen AS. Height and calories in early childhood. Economics & Human Biology. 2016;20:55-69.
64. Bhutta ZA, Ahmed T, Black RE, Cousens S, Dewey K, Giugliani E, et al. What works? Interventions for maternal and child undernutrition and survival. The Lancet.371(9610):417-40.
65. Kennedy GE. From the ape's dilemma to the weanling's dilemma: early weaning and its evolutionary context. Journal of human evolution. 2005;48(2):123-45.
66. Brennan L, McDonald J, Shlomowitz R. Infant feeding practices and chronic child malnutrition in the Indian states of Karnataka and Uttar Pradesh. Economics and human biology. 2004;2(1):139-58.
67. Victora CG, Bahl R, Barros AJD, França GVA, Horton S, Krasevec J, et al. Breastfeeding in the 21st century: epidemiology, mechanisms, and lifelong effect. The Lancet. 2016;387(10017):475-90.
68. Shrimpton R, Victora CG, de Onis M, Lima RC, Blössner M, Clugston G. Worldwide Timing of Growth Faltering: Implications for Nutritional Interventions. Pediatrics. 2001;107(5):e75-e.
69. Victora CG, de Onis M, Hallal PC, Blössner M, Shrimpton R. Worldwide Timing of Growth Faltering: Revisiting Implications for Interventions. Pediatrics. 2010;125(3):e473-e80.
70. Ickes SB, Hurst TE, Flax VL. Maternal Literacy, Facility Birth, and Education Are Positively Associated with Better Infant and Young Child Feeding Practices and Nutritional Status among Ugandan Children. The Journal of Nutrition. 2015.
71. Kipp AM, Blevins M, Haley CA, Mwinga K, Habimana P, Shepherd BE, et al. Factors associated with declining under-five mortality rates from 2000 to 2013: an ecological analysis of 46 African countries. BMJ Open. 2016;6(1).
72. Rutstein SO, and Sarah Staveteig. . Making the Demographic and Health Surveys Wealth Index Comparable. . DHS Methodological Reports. Rockville, Maryland, USA: ICF International. , 2014 Contract No.: 9.